(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 April 2003 (03.04.2003)

PCT

(10) International Publication Number
WO 03/027249 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/30475

(22) International Filing Date: 26 September 2002 (26.09.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/325,277	26 September 2001 (26.09.2001)	US
60/370,526	4 April 2002 (04.04.2002)	US
60/370,620	4 April 2002 (04.04.2002)	US

(71) Applicant (*for all designated States except US*): **SYNGENTA PARTICIPATIONS AG** [CH/CH]; Schwarzwaldallee 215, CH-4058 Basel (CH).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **SU, Wenpei** [CN/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **ANDON, Nancy** [US/US]; Torrey Mesa Research Institute, 3115 Merryfield Row CA 92007, San Diego, CA 92121 (US). **HAYNES, Paul** [GB/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **BRIGGS, Steven, P.** [US/US]; Torrey Mesa Research Institute, 3115 Merry field Row, San Diego, CA 92121 (US). **COOPER, Bret** [US/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **GLAZEBROOK, Jane** [US/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **GOFF, Stephen, A.** [US/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **KATAGIRI, Fumiaki** [JP/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **KREPS, Joel** [US/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **MOUGHAMER, Todd** [US/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **PROVART, Nicholas** [CA/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **RICKE, Darrell** [US/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **ZHU, Tong** [US/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US).

(74) Agent: **TORCHIA, Timothy, E.**; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/027249 A2

(54) Title: HIGH-PROTEIN-PHENOTYPE-ASSOCIATED PLANT GENES

(57) **Abstract:** Methods to identify genes associated with a high protein phenotype in plants are provided, as well as the gene identified thereby and their use.

HIGH-PROTEIN-PHENOTYPE-ASSOCIATED PLANT GENES

Cross-Reference to Related Applications

This application claims the benefit of U.S. Provisional Application No. 60/325,277, filed September 26, 2001, U.S. Provisional Application No. 60/370,526 filed April 4, 2002, and U.S. Provisional Application No. 60/370,620 filed April 4, 2002, each of which is incorporated herein by reference in its entirety.

Reference to Material Submitted on Compact Disc

The sequence listing accompanying this application is contained on compact disc. The material on the CD-ROM (filed herewith), on CD volumes labeled "COPY 1 – SEQUENCE LISTING", "COPY 2 – SEQUENCE LISTING", "COPY 3 – SEQUENCE LISTING" and "CRF", each containing a text file named "60011-PCT Seq List.txt" created September 26, 2002, having a size of 133,120 bytes, is hereby incorporated by reference in its entirety pursuant to 37 C.F.R. § 1.52(e)(5).

Field of the Invention

The present invention generally relates to the field of plant molecular biology, and more specifically to plant genes useful to alter the protein content or level in plants and to develop molecular markers for plant breeding..

Background of the Invention

Farmers grow conventional maize on an estimated 100 million hectares (200 million acres) throughout the developing world. Maize is the world's most widely grown cereal crop and an essential food source for millions of the world's poor. More than half of the world's malnourished children live in countries where maize is an important food. In 20 developing countries, primarily in Latin America and Africa, maize gruel is the main food mothers use to wean their babies, and maize is the single largest source of calories. But babies who subsist on maize can face a dangerous lack of protein during a critical stage of physical and mental development as diets high in maize lack two essential amino acids needed to prevent malnutrition.

In maize crops, the expression of storage protein genes directly affects the nutritional quality of the seed protein. The prolamine (zein) fraction of storage proteins comprises over

50% of the total protein in the mature seed, however, α-zein polypeptides which are especially abundant contain extremely low levels of the essential amino acids lysine and tryptophan. Thus, maize seed protein is deficient in these amino acids because such a large proportion of the total seed storage protein is contributed by the α-zeins (Mertz et al., 1964).

The development of breeding steps to improve maize based on the manipulation of zein profile is hampered by the complexity of the zein proteins. The term "zein" encompasses a family of some 100 related proteins. Zeins can be divided into four structurally distinct types: α-zeins include proteins with molecular weights of 19,000 and 22,000 daltons; β-zeins include proteins with a molecular weight of 14,000 daltons; (γzeins include proteins with molecular weights of 27,000 and 26,000 daltons; and δ-zeins include proteins having a molecular weight of 10,000 daltons. The α-zeins are the major zein proteins found in the endosperm of maize kernels. However, the complexity of zein proteins goes beyond these size classes. Protein sequence analyses indicates that there is microheterogenicity in zein amino acid sequences. This is in accord with isoelectric focusing analyses which show charge differences in zein proteins. Over 70 genes encoding the zein proteins have been identified (Rubenstein, 1982), and the zein genes appear to be located on at least three chromosomes. Thus, the zein proteins are encoded by a multigene family.

There are several mutations known to cause reductions in zein synthesis that lead to alterations in the amino acid content of the seed. For example, in the seeds of plants homozygous for the recessive mutation *opaque-2*, the zein content is reduced by approximately 50% (Tsai et al., 1978). The *opaque-2* mutation primarily affects synthesis of the 19 and 22 kD α-zein proteins, causing a significant decrease in the level of the 19 kD zein fraction and reducing the accumulation of the 22 kD zein fraction to barely detectable levels (Jones et al., 1977). In this mutant, there is a concomitant increase in the proportion of more nutritionally balanced proteins , e.g., albumins, globulins and glutelins, deposited in the seed. The net result of the altered storage protein patterns is an increase in the essential amino acids lysine and tryptophan in the mutant seed (Misra et al., 1972). However, *opaque-2* maize has low yields, chalky-looking grain, and susceptibility to pests and diseases.

Two other recessive mutations, *floury*-2 and *sugary*-1, result in increased levels of methionine in the seed. The increased methionine content in the seeds of *floury*-2 mutants is the result of a decrease in the zein/glutelin ratio, due to reductions in the levels of both the 19 and 22 kD ∀-zein fractions, and an apparent increase in the methionine content of the glutelin fraction (Hansel et al., 1973; Jones, 1978). In *sugary*-1 mutants, there is a decrease in zein

synthesis coupled with an increase in the methionine content of the zein and glutelin fractions (Paulis et al., 1978).

As demonstrated by the *opaque*-2, *floury*-2, and *sugary*-1 mutations, reductions in zein synthesis and/or changes in the relative proportions of the storage protein fractions can affect the overall amino acid composition of the seed. Unfortunately, poor agronomic characteristics (kernel softness, reduced yield, lowered resistance to disease) are associated with the *opaque* and *floury* mutations, preventing their ready application in commercial breeding.

Another way that genes can be down regulated in animals and plants involves the expression of antisense genes. A review of the use of antisense genes in manipulating gene expression in plants can be found in van der Krol et al. (1988a; 1988b). The inhibition of expression of several endogenous plant genes has been reported. For example, U.S. Patent No. 5,107,065 discloses down regulation of polygalacturonase activity by expression of an antisense gene. Other plant genes down regulated using antisense genes include the genes encoding chalcone synthase and the small subunit of ribulose-1,5-biphosphate carboxylase (van der Krol et al., 1988c; Rodermel et al., 1988).

Down regulation of gene expression in a plant may also occur through expression of a particular transgene. This type of down regulation is referred to as co-suppression and involves coordinate silencing of a transgene and a second transgene or a homologous endogenous gene (Matzke and Matzke, 1995). For example, cosuppression of a herbicide resistance gene in tobacco (Brandle et al., 1995), polygalacturonidase in tomato (Flavell, 1994) and chalcone synthase in petunia (U. S. Patent No. 5,034,323) have been demonstrated. Flavell (1994) suggested that multicopy genes, or gene families, must have evolved to avoid cosuppression in order for multiple copies of related genes to be expressed in a plant.

Recently, a new corn variety was prepared which contains nearly twice as much usable protein as other maize grown in the tropics and yields 10 percent more grain. The new maize variety, called "quality protein maize" (QPM), was developed through traditional plant breeding and looks and tastes like normal maize, but the nutritive value of its protein is nearly equivalent to cow's milk. In particular, the varieties produce 70-100 percent more of lysine and tryptophan. A bumper crop of the maize is expected in the coming months from more than one million hectares (2.5 million acres) currently under cultivation in 11 countries. Economists expect that by 2003, the number of hectares sown to QPM will triple to approximately 3.5 million hectares (8.75 million acres). Moreover, as incomes rise in Asia, researchers expect that the use of maize in animal feed will increase by more than three percent

each year between now and the year 2020. The high protein maize fattens pigs and poultry more efficiently, enabling poor farming families to increase their incomes. Pigs and poultry raised on this type of maize gain weight roughly twice as fast as animals fed on conventional maize. However, QPM is the result of more than three decades of scientific discovery.

Thus, there is a need for improved methods to alter the nutritional content of seeds and plants to produce kernels and plants with good agronomic characteristics, while maintaining the phenotype of the parent, e.g., kernel hardness, yield, and disease resistance.

Plants are increasingly used a "protein factories" for production of industrial or therapeutic polypeptides, such as antigens, antibodies (e.g., monoclonal), cytokines, vaccines. Methods for increased yield and/or quality or ease of downstream processing are needed.

Thus, there is a need for improved methods and compositons to alter the protein content of seeds and plants to produce kernels and plants with good characteristics for production of important polypeptides.

Summary of the Invention

Proteins and genes involved in a tropical high protein trait corn germplasm are disclosed, as well as their use to genetically modify cereals for higher protein yield and better protein quality. A total 11 genes (and thir orthologs) are identified for use in protein trait modification in cereals, particularly corn. These genes belong to two groups: one group of proteins is associated with seed protein storage and the other group is generally related to seed stress response or proteins that are unregulated during seed maturation. Possibly the stress response mechanism has co-evolved with the high protein trait. Higher protein yield in corn and other cereals can be achieved by manipulating the gene expression level of these genes and other regulatory genes regulating the stress mechanism.

Accordingly, the invention provides isolated nucleic acid molecules, e.g., DNA, comprising a plant nucleotide sequence encoding a polypeptide that is expressed in cells of a plant, e.g., embryos, mature embryos, endosperm, shoot, root, leaf and developing seed, from high protein varieties of plants, relative to cells of a plant from a corresponding lower protein variety. For example, the invention provides a nucleic acid molecule comprising a plant nucleotide sequence comprising an open reading frame encoding a polypeptide which is substantially similar to a polypeptide comprising any one of SEQ ID Nos. 1-36. To provide altered protein content to a

plant, this sequence may be overexpressed individually, in the sense or antisense orientation, or in combination with other sequences, to confer altered nutritional properties to the plant relative to a plant that does not comprise and/or express the sequence(s). Thus, in one embodiment, the protein content may be enhanced, while in another embodiment it may be reduced, e.g., low protein products such as rice for individuals that are intolerant or sensitive to certain proteins. Low protein content plants or seeds can be a superior form for production of heterologous industrially or therapeutically important proteins in plants, and plant seeds by, for example, reducing levels of abundant endogenous proteins. To avoid detrimental effects to the plants, such modulation can be controlled using inducible promoters. One system employs hybrid two component systems such as Gal4/C1, in which the controlled promoter(s) is on or off only in the hybrid, not the parental lines. The overexpression may be constitutive, or it may be preferable to express the sequence from an inducible promoter including a promoter which is responsive to external stimuli, such as chemical application, or environmental stimuli, so as to avoid possible deleterious effects on plant growth. High protein varieties of plants are those which have at least a 1%, preferably at least 5%, and more preferably at least 10%, increase in protein content or level relative to a corresponding control plant. For example, for maize, a high protein line or variety preferably may have a protein content in whole kernel that is at least 14.5%, more preferably at least 15.5%, in embyro at least 17%, more preferably 18.3%, and in endosperm at least 13.5%, more preferably at least 14.2%. High protein varieties of maize are well known to the art, see, for example, U.S. Patent Nos. 5,986,182, 5,936,143, 5,907,089, 5,900,528, 5,850,031, 5,824,855, 5,824,854, 5,763,756, and 5,675,065.

As described herein, protein expression profiles from embryos of normal and high protein varieties of maize were compared using two-dimensional SDS-PAGE analysis in order to identify differentially expressed genes. Application of proteomic technology to the high protein corn germplasm has revealed more than 120 genes that are differentially expressed in high protein lines. Such genes may encode structural or regulatory proteins, and hence are of potential use in manipulating protein content in maize (corn) and other cereals such as wheat and rice, e.g., for manipulating seed protein phenotype and for the development of molecular markers for plant breeding. Moreover, based on the proteomic approach, the results provide a novel function for

unknown or previously uncharacterized/ mischaracterized genes, and may lead to useful regulatory genes for particular traits, structural genes or molecular markers. Further, by using a segregating population, the results also provide the necessary means to identify genes specifically related to the high protein phenotype rather than those that are merely causally associated. Thus, the identified proteins (polypeptides) and their corresponding genes can be used to: 1) manipulate protein content or levels in corn and other cereal species, e.g., by using the genes as molecular markers in breeding or in transgenic plants; 2) isolate orthologs from other crop species such as rice and wheat; 3) generate antibodies and develop protein-based assays for breeding selection; and 4) identify common transcriptional regulatory elements and factors which bind those elements, i.e., the upstream regions of the genes associated with the high protein trait.

Non-protein based methods may also be employed to identify the nucleic acid molecules of the invention. For example, an array of nucleic acid samples, e.g., a plurality of oligonucleotides, each plurality corresponding to a different plant gene, on a solid substrate, e.g., a DNA chip, and probes corresponding to nucleic acid obtained from plant sources that express genes associated with protein content and probes to nucleic acid obtained from plant sources that do not express those genes or express the genes at a reduced level, can be used to systematically identify genes associated with increased protein levels.

Preferably, the nucleotide sequence in the nucleic acid molecule of the invention is from plant DNA, either a dicot or a monocot, which encodes a polypeptide that is substantially similar to a polypeptide comprising any one of SEQ ID NOs: 1-36. More preferably, the nucleotide sequence is from plant DNA that is substantially similar to a nucleic acid segment encoding a polypeptide comprising any one of SEQ ID NOs: 1-36. The term "substantially similar", when used herein with respect to a polypeptide means a polypeptide corresponding to a reference polypeptide, wherein the polypeptide has substantially the same structure and function as the reference polypeptide, e.g., where only changes in amino acid sequence are those which do not affect the polypeptide function. When used for a polypeptide or an amino acid sequence, the percentage of identity between the substantially similar and the reference polypeptide or amino acid sequence is at least 65%, 66%, 67%, 68%, 69%, 70%, e.g., 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, and even 90% or more, e.g., 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, up to at least 99%, where the reference polypeptide is a polypeptide comprising any one of SEQ ID NOs: 1-36. One indication that two polypeptides are substantially similar to each other is that

an agent, e.g., an antibody, which specifically binds to one of the polypeptides, specifically binds to the other.

In its broadest sense, the term "substantially similar", when used herein with respect to a nucleotide sequence, means a nucleotide sequence corresponding to a reference nucleotide sequence, wherein the corresponding sequence encodes a polypeptide having substantially the same structure and function as the polypeptide encoded by the reference nucleotide sequence. The term "substantially similar" is specifically intended to include nucleotide sequences wherein the sequence has been modified to optimize expression in particular cells. The percentage of identity between the substantially similar nucleotide sequence and the reference nucleotide sequence is at least 65%, 66%, 67%, 68%, 69%, 70%, e.g., 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, and even 90% or more, e.g., 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, up to at least 99%, wherein the reference sequence is one which encodes a polypeptide comprising any one of SEQ ID NOs: 1-36, or the complement thereof. Sequence comparisons maybe carried out using a Smith-Waterman sequence alignment algorithm (see e.g. Waterman (1995) or http://www hto.usc.edu/software/seqaln/index.html). The localS program, version 1.16, is preferably used with following parameters: match: 1, mismatch penalty: 0.33, open-gap penalty: 2, extended-gap penalty: 2. Further, a nucleotide sequence that is "substantially similar" to a reference nucleotide sequence hybridizes to the reference nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C, more desirably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1% SDS at 50°C, more desirably still in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C, preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C, more preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C.

Hence, the isolated nucleic acid molecules of the invention also include orthologs of the sequences encoding the polypeptides comprising the amino acid sequences disclosed herein, including, but not limited to, dicots and monocots, preferably cereal plants, e.g., wheat or rice. An ortholog is a gene from a different species that encodes a product having the same function as the product encoded by a gene from a reference organism. The encoded ortholog products likely have at least 70% sequence identity to each other. Hence, the invention

includes an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide having at least 70% identity to a polypeptide comprising one or more of the sequences disclosed herein. Databases such GenBank or one found at http://bioserver.myongjiac.kr/rjce.html (for rice) may be employed to identify sequences related to the disclosed sequences, e.g., orthologs in cereal crops such as rice. Alternatively, recombinant DNA techniques such as hybridization or PCR may be employed to identify sequences related to the disclosed sequences.

Preferably, the polypeptide has substantial identity, i.e., at least 70%, e.g., 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, and even 90% or more, e.g., 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, and at least 99%, amino acid sequence identity to a polypeptide comprising any one of SEQ ID NOs: 1-36. The invention also provides anti-sense nucleic acid molecules corresponding to the genes identified herein. Also provided are expression cassettes, e.g., recombinant vectors, and host cells, comprising the nucleic acid molecule of the invention.

The nucleic acid molecules of the invention, their encoded polypeptides and compositions thereof, are useful to provide plants with enhanced protein content, identify common transcriptional regulatory factors which bind upstream of the coding region of genes associated with high protein content and as markers for breeding selection. The compositions of the invention include plant nucleic acid sequences and the amino acid sequences for the polypeptides or partial-length polypeptides encoded thereby which are useful to provide enhanced nutritional characteristics to a plant, preferably by enhancing protein content or levels. Methods of the invention involve stably transforming a plant with one or more of at least a portion of these nucleotide sequences operably linked to a promoter capable of driving expression of that nucleotide sequence in a plant cell. By "portion" or "fragment", as it relates to a nucleic acid molecule, sequence or segment of the invention, when it is linked to other sequences for expression, is meant a sequence having at least 80 nucleotides, more preferably at least 150 nucleotides, and still more preferably at least 400 nucleotides. If not employed for expressing, a "portion" or "fragment" means at least 9, preferably 12, more preferably 15, even more preferably at least 20, consecutive nucleotides, e.g., probes and primers (oligonucleotides), corresponding to the nucleotide sequence of the nucleic acid molecules of the invention. The method comprises introducing to a plant, plant cell, or plant tissue an expression cassette comprising at least one of nucleic acid molecules of the invention so as to yield a transformed differentiated plant, transformed cell or transformed tissue. Transformed

cells or tissue can be regenerated to provide a transformed differentiated plant. The transformed differentiated plant preferably expresses the nucleic acid molecule in an amount that yields a transformed plant having enhanced protein content, e.g., in seed, to a corresponding nontransformed plant. The present invention also provides a transformed plant prepared by the method, progeny and seed thereof.

A transformed (transgenic) plant of the invention includes plants, dicots or monocots, the genome of which is augmented by a nucleic acid molecule of the invention, or in which the corresponding gene has been disrupted, e.g., to result in a loss, a decrease or an alteration, in the function of the product encoded by the gene. The nucleic acid molecules of the invention are thus useful for targeted gene disruption, as well as for markers and probes.

The invention also includes recombinant nucleic acid molecules which have been modified so as to comprise codons other than those present in the unmodified sequence. The recombinant nucleic acid molecules of the invention include those in which the modified codons specify amino acids that are the same as those specified by the codons in the unmodified sequence, as well as those that specify different amino acids, i.e., they encode a variant polypeptide having one or more amino acid substitutions relative to the polypeptide encoded by the unmodified sequence.

The invention further includes a nucleotide sequence which is complementary to one (hereinafter "test" sequence) which hybridizes under stringent conditions with the nucleic acid molecules of the invention as well as RNA which is encoded by the nucleic acid molecule. When the hybridization is performed under stringent conditions, either the test or nucleic acid molecule of invention is preferably supported, e.g., on a membrane or DNA chip. Thus, either a denatured test or nucleic acid molecule of the invention is preferably first bound to a support and hybridization is effected for a specified period of time at a temperature of, e.g., between 55 and 70°C, in double strength citrate buffered saline (SC) containing 0.1% SDS followed by rinsing of the support at the same temperature but with a buffer having a reduced SC concentration. Depending upon the degree of stringency required such reduced concentration buffers are typically single strength SC containing 0.1% SDS, half strength SC containing 0.1% SDS and one-tenth strength SC containing 0.1% SDS.

The present invention also provides a method to identify a polypeptide which is associated with a high protein phenotype. The method comprises separating a plurality of polypeptides from a sample comprising polypeptides, wherein the sample is from a plant having a high protein content. Then the separated sample of polypeptides from a plant having

a high protein content is compared to a separated sample of polypeptides from a corresponding plant with lower protein content. Preferably, polypeptides are identified that are present in the sample from a plant having a high protein content that are not present in the sample from the plant with lower protein content.

Also provided is an isolated nucleic acid molecule comprising a nucleotide sequence that directs transcription, e.g., a promoter, of a linked nucleic acid fragment in a host cell, such as a plant cell. It is preferred that the nucleotide sequence is from plant genomic DNA which has at least 65%, 66%, 67%, 68%, 69%, 70%, e.g., 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, and even 90% or more, e.g., 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, and 99%, nucleotide sequence identity to a sequence of a promoter from a plant gene encoding a polypeptide comprising any one of SEQ ID NOs: 1-36. The promoter sequence is preferably about 25 to 2000, e.g., 50 to 500 or 100 to 1400, nucleotides in length. In one embodiment of the invention, the isolated nucleic acid molecule comprises a plant nucleotide sequence which is the promoter region for a gene encoding any one of SEQ ID NOs: 1-36, or is structurally related to the promoter for a gene encoding SEQ ID NOs: 1-36, i.e., is an orthologous promoter, and is linked to a plant structural gene. Hence, the present invention further provides an expression cassette or a recombinant vector containing the nucleic acid molecule, and the vector may be a plasmid. Such cassettes or vectors, when present in a plant, plant cell or plant tissue result in transcription of the linked nucleic acid fragment in the plant, plant tissue or plant cell.

The expression cassettes or vectors of the invention may optionally include other regulatory sequences, e.g., transcription terminator sequences, introns and/or enhancers, and may be contained in a host cell. The expression cassette or vector may augment the genome of a transformed plant or may be maintained extrachromosomally. The expression cassette or vector may further have a Ti plasmid and be contained in an *Agrobacterium tumefaciens* cell; it may be carried on a microparticle, wherein the microparticle is suitable for ballistic transformation of a plant cell; or it may be contained in a plant cell protoplast. Further, the expression cassette can be contained in a plant, plant cell or plant tissue from a dicot or a monocot. In particular, the plant may be a cereal plant.

The present invention further provides a method of augmenting a plant genome by contacting plant cells with an expression cassette or vector of the invention, i.e., one having a nucleotide sequence that directs transcription of a linked nucleic acid fragment in a plant cell, wherein the nucleotide sequence is from plant genomic DNA that has at least 65%, and more

preferably at least 70%, identity to the sequence of a promoter from a gene encoding a polypeptide comprising any one of SEQ ID NOs: 1-36 so as to yield transformed plant cells; and regenerating the transformed plant cells to provide a differentiated transformed plant, wherein the differentiated transformed plant expresses the linked fragment in the cells of the plant. The present invention also provides a plant prepared by the method, progeny and seed thereof.

Brief Description of the Figures

Figure 1 shows the protein content in various sources from high protein and control maize lines.

Figures 2A and 2B illustrate a two dimensional gel with proteins from a control (#530; panel A) or high protein (#465; panel B) maize line. Figures 2C and 2D illustrate another comparision of protein expression profile of high protein germplasm and normal corn line.

Figures 3A to 3H show the peptide and criteria (e.g., Xcro > 2 and Dcn > 0.01) employed to search databases for the corresponding full length protein for 18 of the proteins shown in the attached Sequence Listing which is incorporated herein.

Figures 4A and 4B are representative vectors for over- or under-expression of genes in seed.

The Sequence Listing shows the amino acid sequence of proteins, high protein phenotype genes and proteins, or the orthologs thereof, which are preferentially expressed in high protein maize lines relative to lines with lower protein content. Sequences 1 to 36 are the high protein involved proteins. Odd numbered SEQ ID Nos are are protein-encoding orfs and the even numbered SEQ ID NOs are amino acids sequences. Sequences 37 to 45 are representative peptides identified by the MS as described in the Examples.

Detailed Description of the Invention

Definitions

The term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form, composed of monomers (nucleotides) containing a sugar, phosphate and a base which is either a purine or pyrimidine. Unless specifically limited, the term encompasses nucleic acids containing known analogs of natural nucleotides which have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively

modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer et al., 1991; Ohtsuka et al., 1985; Rossolini et al., 1994). A "nucleic acid fragment" is a fraction of a given nucleic acid molecule. In higher plants, deoxyribonucleic acid (DNA) is the genetic material while ribonucleic acid (RNA) is involved in the transfer of information contained within DNA into proteins. The term "nucleotide sequence" refers to a polymer of DNA or RNA which can be single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases capable of incorporation into DNA or RNA polymers. The terms "nucleic acid", "nucleotide sequence", "nucleic acid molecule", "nucleic acid fragment" or "nucleic acid sequence or segment" may also be used interchangeably with gene, cDNA, DNA and RNA encoded by a gene.

The invention encompasses isolated or substantially purified nucleic acid or protein compositions. In the context of the present invention, an "isolated" or "purified" DNA molecule or an "isolated" or "purified" polypeptide is a DNA molecule or polypeptide that, by the hand of man, exists apart from its native environment and is therefore not a product of nature. An isolated DNA molecule or polypeptide may exist in a purified form or may exist in a non-native environment such as, for example, a transgenic host cell. For example, an "isolated" or "purified" nucleic acid molecule or protein, or biologically active portion thereof, is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. In one embodiment, an "isolated" nucleic acid is free of sequences that naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, or 0.1 kb of nucleotide sequences that naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. A protein that is substantially free of cellular material includes preparations of protein or polypeptide having less than about 30%, 20%, 10%, 5%, (by dry weight) of contaminating protein. When the protein of the invention, or biologically active portion thereof, is recombinantly produced, preferably culture medium represents less than about 30%, 20%, 10%, or 5% (by dry weight) of chemical precursors or non-protein-of- interest chemicals.

Fragments and variants of the disclosed nucleotide sequences and proteins or partial-length proteins encoded thereby are also encompassed by the present invention. By "fragment" or "portion" is meant a full length or less than full length of the nucleotide sequence encoding, or the amino acid sequence of, a polypeptide or protein. Alternatively, fragments or portions of a nucleotide sequence that are useful as hybridization probes generally do not encode fragment proteins retaining biological activity. Thus, fragments or portions of a nucleotide sequence may range from at least about 9 nucleotides, about 12 nucleotides, about 20 nucleotides, about 50 nucleotides, about 100 nucleotides or more.

The term "gene" is used broadly to refer to any segment of nucleic acid associated with a biological function. Thus, genes include coding sequences and/or the regulatory sequences required for their expression. For example, gene refers to a nucleic acid fragment that expresses mRNA, functional RNA, or specific protein, including regulatory sequences. Genes also include nonexpressed DNA segments that, for example, form recognition sequences for other proteins. Genes can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters.

"Naturally occurring" is used to describe an object that can be found in nature as distinct from being artificially produced by man. For example, a protein or nucleotide sequence present in an organism (including a virus), which can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory, is naturally occurring.

A "marker gene" encodes a selectable or screenable trait.

“Selectable marker” is a gene whose expression in a cell gives the cell a selective advantage. The selective advantage possessed by the cells transformed with the selectable marker gene may be due to their ability to grow in the presence of a negative selective agent, such as an antibiotic or a herbicide, compared to the growth of non-transformed cells. The selective advantage possessed by the transformed cells, compared to non-transformed cells, may also be due to their enhanced or novel capacity to utilize an added compound as a nutrient, growth factor or energy source. Selectable marker gene also refers to a gene or a combination of genes whose expression in a cell gives the cell both a negative and/or a positive selective advantage.

The term "chimeric" refers to any gene or DNA that contains 1) DNA sequences, including regulatory and coding sequences, that are not found together in nature, or 2)

sequences encoding parts of proteins not naturally adjoined, or 3) parts of promoters that are not naturally adjoined. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or comprise regulatory sequences and coding sequences derived from the same source, but arranged in a manner different from that found in nature.

A "transgene" refers to a gene that has been introduced into the genome by transformation and is stably maintained. Transgenes may include, for example, DNA that is either heterologous or homologous to the DNA of a particular plant to be transformed. Additionally, transgenes may comprise native genes inserted into a non-native organism, or chimeric genes. The term "endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism but that is introduced by gene transfer.

The terms "protein," "peptide" and "polypeptide" may be used interchangeably herein.

By "variants" is intended substantially similar sequences. For nucleotide sequences, variants include those sequences that, because of the degeneracy of the genetic code, encode the identical amino acid sequence of the native protein. Naturally occurring allelic variants such as these can be identified with the use of well-known molecular biology techniques, as, for example, with polymerase chain reaction (PCR) and hybridization techniques. Variant nucleotide sequences also include synthetically derived nucleotide sequences, such as those generated, for example, by using site-directed mutagenesis which encode the native protein, as well as those that encode a polypeptide having amino acid substitutions. Generally, nucleotide sequence variants of the invention will have at least 40, 50, 60, to 70%, e.g., preferably 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, to 79%, generally at least 80%, e.g., 81%-84%, at least 85%, e.g., 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, to 98%, sequence identity to the native (endogenous) nucleotide sequence.

“DNA shuffling” is a method to introduce mutations or rearrangements, preferably randomly, in a DNA molecule or to generate exchanges of DNA sequences between two or more DNA molecules, preferably randomly. The DNA molecule resulting from DNA shuffling is a shuffled DNA molecule that is a non-naturally occurring DNA molecule derived from at least one template DNA molecule. The shuffled DNA preferably encodes a variant polypeptide modified with respect to the polypeptide encoded by the template DNA, and may have an altered biological activity with respect to the polypeptide encoded by the template DNA.

The nucleic acid molecules of the invention can be optimized for enhanced expression in plants of interest. See, for example, EPA035472; WO91/16432; Perlak et al., 1991; and Murray et al., 1989. In this manner, the genes or gene fragments can be synthesized utilizing plant-preferred codons. See, for example, Campbell and Gowri (1990) for a discussion of host-preferred codon usage. Thus, the nucleotide sequences can be optimized for expression in any plant. It is recognized that all or any part of the gene sequence may be optimized or synthetic. That is, synthetic or partially optimized sequences may also be used. Variant nucleotide sequences and proteins also encompass sequences and protein derived from a mutagenic and recombinogenic procedure such as DNA shuffling. With such a procedure, one or more different coding sequences can be manipulated to create a new polypeptide possessing the desired properties. In this manner, libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides comprising sequence regions that have substantial sequence identity and can be homologously recombined *in vitro* or *in vivo*. Strategies for such DNA shuffling are known in the art. See, for example, Stemmer (1994); Stemmer (1994); Crameri et al. (1997); Moore et al. (1997); Zhang et al. (1997); Crameri et al. (1998); and U.S. Patent Nos. 5,605,793 and 5,837,458.

"Conservatively modified variations" of a particular nucleic acid sequence refers to those nucleic acid sequences that encode identical or essentially identical amino acid sequences, or where the nucleic acid sequence does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given polypeptide. For instance the codons CGT, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at every position where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded protein. Such nucleic acid variations are "silent variations" which are one species of "conservatively modified variations." Every nucleic acid sequence described herein which encodes a polypeptide also describes every possible silent variation, except where otherwise noted. One of skill will recognize that each codon in a nucleic acid (except ATG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule by standard techniques. Accordingly, each "silent variation" of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

"Recombinant DNA molecule" is a combination of DNA sequences that are joined together using recombinant DNA technology and procedures used to join together DNA sequences as described, for example, in Sambrook et al.

The terms "heterologous DNA sequence," "exogenous DNA segment" or "heterologous nucleic acid," each refer to a sequence that originates from a source foreign to the particular host cell or, if from the same source, is modified from its original form. Thus, a heterologous gene in a host cell includes a gene that is endogenous to the particular host cell but has been modified through, for example, the use of DNA shuffling. The terms also include non-naturally occurring multiple copies of a naturally occurring DNA sequence. Thus, the terms refer to a DNA segment that is foreign or heterologous to the cell, or homologous to the cell but in a position within the host cell nucleic acid in which the element is not ordinarily found. Exogenous DNA segments are expressed to yield exogenous polypeptides.

A "homologous" DNA sequence is a DNA sequence that is naturally associated with a host cell into which it is introduced.

"Wild-type" refers to the normal gene, or organism found in nature without any known mutation.

"Genome" refers to the complete genetic material of an organism.

"Vector" is defined to include, inter alia, any plasmid, cosmid, phage or *Agrobacterium* binary vector in double or single stranded linear or circular form which may or may not be self transmissible or mobilizable, and which can transform prokaryotic or eukaryotic host either by integration into the cellular genome or exist extrachromosomally (e.g., autonomous replicating plasmid with an origin of replication).

Specifically included are shuttle vectors by which is meant a DNA vehicle capable, naturally or by design, of replication in two different host organisms, which may be selected from actinomycetes and related species, bacteria and eukaryotic (e.g., higher plant, mammalian, yeast or fungal cells).

"Cloning vectors" typically contain one or a small number of restriction endonuclease recognition sites at which foreign DNA sequences can be inserted in a determinable fashion without loss of essential biological function of the vector, as well as a marker gene that is suitable for use in the identification and selection of cells transformed with the cloning vector. Marker genes typically include genes that provide tetracycline resistance, hygromycin resistance or ampicillin resistance.

"Expression cassette" as used herein means a DNA sequence capable of directing expression of a particular nucleotide sequence in an appropriate host cell, comprising a promoter operably linked to the nucleotide sequence of interest which is operably linked to termination signals. It also typically comprises sequences required for proper translation of the nucleotide sequence. The coding region usually codes for a protein of interest but may also code for a functional RNA of interest, for example antisense RNA or a nontranslated RNA, in the sense or antisense direction. The expression cassette comprising the nucleotide sequence of interest may be chimeric, meaning that at least one of its components is heterologous with respect to at least one of its other components. The expression cassette may also be one which is naturally occurring but has been obtained in a recombinant form useful for heterologous expression. The expression of the nucleotide sequence in the expression cassette may be under the control of a constitutive promoter or of an inducible promoter which initiates transcription only when the host cell is exposed to some particular external stimulus. In the case of a multicellular organism, the promoter can also be specific to a particular tissue or organ or stage of development.

Such expression cassettes will comprise the transcriptional initiation region of the invention linked to a nucleotide sequence of interest. Such an expression cassette is provided with a plurality of restriction sites for insertion of the gene of interest to be under the transcriptional regulation of the regulatory regions. The expression cassette may additionally contain selectable marker genes.

The transcriptional cassette will include in the 5'-3' direction of transcription, a transcriptional and translational initiation region, a DNA sequence of interest, and a transcriptional and translational termination region functional in plants. The termination region may be native with the transcriptional initiation region, may be native with the DNA sequence of interest, or may be derived from another source. Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. See also, Guerineau et al. (1991); Proudfoot (1991); Sanfacon et al. (1991); Mogen et al. (1990); Munroe et al. (1990); Ballas et al. (1989); Joshi et al. (1987).

An oligonucleotide corresponding to a nucleic acid molecule of the invention may be about 30 or fewer nucleotides in length (e.g., 9, 12, 15, 18, 20, 21 or 24, or any number between 9 and 30). Generally specific primers are upwards of 14 nucleotides in length. For optimum specificity and cost effectiveness, primers of 16-24 nucleotides in length may be

preferred. Those skilled in the art are well versed in the design of primers for use processes such as PCR. If required, probing can be done with entire restriction fragments of the gene disclosed herein which may be 100's or even 1000's of nucleotides in length.

"Coding sequence" refers to a DNA or RNA sequence that codes for a specific amino acid sequence and excludes the non-coding sequences. It may constitute an "uninterrupted coding sequence", i.e., lacking an intron, such as in a cDNA or it may include one or more introns bounded by appropriate splice junctions. An "intron" is a sequence of RNA which is contained in the primary transcript but which is removed through cleavage and re-ligation of the RNA within the cell to create the mature mRNA that can be translated into a protein.

The terms "open reading frame" and "ORF" refer to the amino acid sequence encoded between translation initiation and termination codons of a coding sequence. The terms "initiation codon" and "termination codon" refer to a unit of three adjacent nucleotides ('codon') in a coding sequence that specifies initiation and chain termination, respectively, of protein synthesis (mRNA translation).

A "functional RNA" refers to an antisense RNA, ribozyme, or other RNA that is not translated.

The term "RNA transcript" refers to the product resulting from RNA polymerase catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA" (mRNA) refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a single- or a double-stranded DNA that is complementary to and derived from mRNA.

"Regulatory sequences" and "suitable regulatory sequences" each refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences include enhancers, promoters, translation leader sequences, introns, and polyadenylation signal sequences. They include natural and synthetic sequences as well as sequences which may be a combination of synthetic and natural sequences. As is noted above, the term "suitable regulatory sequences" is not limited to promoters. However, some suitable regulatory sequences useful in the present invention will include, but are not limited to constitutive plant

promoters, plant tissue-specific promoters, plant development specific promoters, inducible plant promoters and viral promoters.

"5' non-coding sequence" refers to a nucleotide sequence located 5' (upstream) to the coding sequence. It is present in the fully processed mRNA upstream of the initiation codon and may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency (Turner et al., 1995).

"3' non-coding sequence" refers to nucleotide sequences located 3' (downstream) to a coding sequence and include polyadenylation signal sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al., 1989.

The term "translation leader sequence" refers to that DNA sequence portion of a gene between the promoter and coding sequence that is transcribed into RNA and is present in the fully processed mRNA upstream (5') of the translation start codon. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency.

The term "mature" protein refers to a post-translationally processed polypeptide without its signal peptide. "Precursor" protein refers to the primary product of translation of an mRNA. "Signal peptide" refers to the amino terminal extension of a polypeptide, which is translated in conjunction with the polypeptide forming a precursor peptide and which is required for its entrance into the secretory pathway. The term "signal sequence" refers to a nucleotide sequence that encodes the signal peptide.

The term "intracellular localization sequence" refers to a nucleotide sequence that encodes an intracellular targeting signal. An "intracellular targeting signal" is an amino acid sequence that is translated in conjunction with a protein and directs it to a particular sub-cellular compartment. "Endoplasmic reticulum (ER) stop transit signal" refers to a carboxy-terminal extension of a polypeptide, which is translated in conjunction with the polypeptide and causes a protein that enters the secretory pathway to be retained in the ER. "ER stop transit sequence" refers to a nucleotide sequence that encodes the ER targeting signal. Other intracellular targeting sequences encode targeting signals active in seeds and/or leaves and vacuolar targeting signals.

"Promoter" refers to a nucleotide sequence, usually upstream (5') to its coding sequence, which controls the expression of the coding sequence by providing the recognition for RNA polymerase and other factors required for proper transcription. "Promoter" includes a minimal promoter that is a short DNA sequence comprised of a TATA- box and other sequences that serve to specify the site of transcription initiation, to which regulatory elements are added for control of expression. "Promoter" also refers to a nucleotide sequence that includes a minimal promoter plus regulatory elements that is capable of controlling the expression of a coding sequence or functional RNA. This type of promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue specificity of a promoter. It is capable of operating in both orientations (normal or flipped), and is capable of functioning even when moved either upstream or downstream from the promoter. Both enhancers and other upstream promoter elements bind sequence-specific DNA-binding proteins that mediate their effects. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even be comprised of synthetic DNA segments. A promoter may also contain DNA sequences that are involved in the binding of protein factors which control the effectiveness of transcription initiation in response to physiological or developmental conditions.

The "initiation site" is the position surrounding the first nucleotide that is part of the transcribed sequence, which is also defined as position +1. With respect to this site all other sequences of the gene and its controlling regions are numbered. Downstream sequences (i.e., further protein encoding sequences in the 3' direction) are denominated positive, while upstream sequences (mostly of the controlling regions in the 5' direction) are denominated negative.

Promoter elements, particularly a TATA element, that are inactive or that have greatly reduced promoter activity in the absence of upstream activation are referred to as "minimal or core promoters." In the presence of a suitable transcription factor, the minimal promoter functions to permit transcription. A "minimal or core promoter" thus consists only of all basal elements needed for transcription initiation, e.g., a TATA box and/or an initiator.

"Constitutive expression" refers to expression using a constitutive or regulated promoter. "Conditional" and "regulated expression" refer to expression controlled by a regulated promoter.

"Constitutive promoter" refers to a promoter that is able to express the gene that it controls in all or nearly all of the plant tissues during all or nearly all developmental stages of the plant. Each of the transcription-activating elements do not exhibit an absolute tissue-specificity, but mediate transcriptional activation in most plant parts at a level of ≥1% of the level reached in the part of the plant in which transcription is most active.

"Regulated promoter" refers to promoters that direct gene expression not constitutively, but in a temporally- and/or spatially-regulated manner, and include both tissue-specific and inducible promoters. It includes natural and synthetic sequences as well as sequences which may be a combination of synthetic and natural sequences. Different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. New promoters of various types useful in plant cells are constantly being discovered, numerous examples may be found in the compilation by Okamuro et al. (1989). Since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity. Typical regulated promoters useful in plants include but are not limited to safener-inducible promoters, promoters derived from the tetracycline-inducible system, promoters derived from salicylate-inducible systems, promoters derived from alcohol-inducible systems, promoters derived from glucocorticoid-inducible system, promoters derived from pathogen-inducible systems, and promoters derived from ecdysome-inducible systems.

"Tissue-specific promoter" refers to regulated promoters that are not expressed in all plant cells but only in one or more cell types in specific organs (such as leaves or seeds), specific tissues (such as embryo or cotyledon), or specific cell types (such as leaf parenchyma or seed storage cells). These also include promoters that are temporally regulated, such as in early or late embryogenesis, during fruit ripening in developing seeds or fruit, in fully differentiated leaf, or at the onset of senescence.

"Inducible promoter" refers to those regulated promoters that can be turned on in one or more cell types by an external stimulus, such as a chemical, light, hormone, stress, or a pathogen.

"Operably-linked" refers to the association of nucleic acid sequences on single nucleic acid fragment so that the function of one is affected by the other. For example, a regulatory

DNA sequence is said to be "operably linked to" or "associated with" a DNA sequence that codes for an RNA or a polypeptide if the two sequences are situated such that the regulatory DNA sequence affects expression of the coding DNA sequence (i.e., that the coding sequence or functional RNA is under the transcriptional control of the promoter). Coding sequences can be operably-linked to regulatory sequences in sense or antisense orientation.

"Expression" refers to the transcription and/or translation of an endogenous gene or a transgene in plants. For example, in the case of antisense constructs, expression may refer to the transcription of the antisense DNA only. In addition, expression refers to the transcription and stable accumulation of sense (mRNA) or functional RNA. Expression may also refer to the production of protein.

"Altered levels" refers to the level of expression in transgenic cells or organisms that differs from that of normal or untransformed cells or organisms.

"Overexpression" refers to the level of expression in transgenic cells or organisms that exceeds levels of expression in normal or untransformed cells or organisms.

"Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of protein from an endogenous gene or a transgene.

"Co-suppression" and "transwitch" each refer to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar transgene or endogenous genes (U.S. Patent No. 5,231,020).

"Gene silencing" refers to homology-dependent suppression of viral genes, transgenes, or endogenous nuclear genes. Gene silencing may be transcriptional, when the suppression is due to decreased transcription of the affected genes, or post-transcriptional, when the suppression is due to increased turnover (degradation) of RNA species homologous to the affected genes. (English et al., 1996). Gene silencing includes virus-induced gene silencing (Ruiz et al., 1998).

"Silencing suppressor" gene refers to a gene whose expression leads to counteracting gene silencing and enhanced expression of silenced genes. Silencing suppressor genes may be of plant, non-plant, or viral origin. Examples include, but are not limited to HC-Pro, P1-HC-Pro, and 2b proteins. Other examples include one or more genes in TGMV-B genome.

"Transcription stop fragment" refers to nucleotide sequences that contain one or more regulatory signals, such as polyadenylation signal sequences, capable of terminating transcription. Examples include the 3' non-regulatory regions of genes encoding nopaline synthase and the small subunit of ribulose bisphosphate carboxylase.

"Translation stop fragment" refers to nucleotide sequences that contain one or more regulatory signals, such as one or more termination codons in all three frames, capable of terminating translation. Insertion of a translation stop fragment adjacent to or near the initiation codon at the 5' end of the coding sequence will result in no translation or improper translation. Excision of the translation stop fragment by site-specific recombination will leave a site-specific sequence in the coding sequence that does not interfere with proper translation using the initiation codon.

The terms "*cis*-acting sequence" and "*cis*-acting element" refer to DNA or RNA sequences whose functions require them to be on the same molecule. An example of a *cis*-acting sequence on the replicon is the viral replication origin.

The terms "*trans*-acting sequence" and "*trans*-acting element" refer to DNA or RNA sequences whose function does not require them to be on the same molecule.

"Chromosomally-integrated" refers to the integration of a foreign gene or DNA construct into the host DNA by covalent bonds. Where genes are not "chromosomally integrated" they may be "transiently expressed." Transient expression of a gene refers to the expression of a gene that is not integrated into the host chromosome but functions independently, either as part of an autonomously replicating plasmid or expression cassette, for example, or as part of another biological system such as a virus.

The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides: (a) "reference sequence", (b) "comparison window", (c) "sequence identity", (d) "percentage of sequence identity", and (e) "substantial identity".

(a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full length cDNA or gene sequence, or the complete cDNA or gene sequence.

(b) As used herein, "comparison window" makes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of sequences for comparison are well known in the art. Thus, the determination of percent identity between any two sequences can be accomplished using a mathematical algorithm. Preferred, non-limiting examples of such mathematical algorithms are the algorithm of Myers and Miller (1988); the local homology algorithm of Smith et al. (1981); the homology alignment algorithm of Needleman and Wunsch (1970); the search-for-similarity-method of Pearson and Lipman (1988); the algorithm of Karlin and Altschul (1990), modified as in Karlin and Altschul (1993).

Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations include, but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, California); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wisconsin, USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins et al. (1988); Higgins et al. (1989); Corpet et al. (1988); Huang et al. (1992); and Pearson et al. (1994). The ALIGN program is based on the algorithm of Myers and Miller, *supra*. The BLAST programs of Altschul et al. (1990), are based on the algorithm of Karlin and Altschul *supra*.

Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., 1990). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity X from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments, or the end of either sequence is reached.

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul (1993). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a test nucleic acid sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid sequence to the reference nucleic acid sequence is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul et al. (1997). Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. See Altschul et al., *supra*. When utilizing BLAST, Gapped BLAST, PSI-BLAST, the default parameters of the respective programs (e.g. BLASTN for nucleotide sequences, BLASTX for proteins) can be used. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, 1989). See http://www.ncbi.nlm.nih.gov. Alignment may also be performed manually by inspection.

For purposes of the present invention, comparison of nucleotide sequences for determination of percent sequence identity to the promoter sequences disclosed herein is preferably made using the BlastN program (version 1.4.7 or later) with its default parameters or any equivalent program. By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by the preferred program.

(c) As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences makes reference to a specified percentage of residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window, as measured by sequence comparison algorithms or by visual inspection. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions,

where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have "sequence similarity" or "similarity." Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California).

(d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.

(e)(i) The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, or 79%, preferably at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%, more preferably at least 90%, 91%, 92%, 93%, or 94%, and most preferably at least 95%, 96%, 97%, 98%, or 99% sequence identity, compared to a reference sequence using one of the alignment programs described using standard parameters. One of skill in the art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning, and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 70%, more preferably at least 80%, 90%, and most preferably at least 95%.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions (see below). Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. However, stringent conditions encompass temperatures in the range of about 1°C to about 20°C, depending upon the desired degree of stringency as otherwise qualified herein. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides they encode are substantially identical. This may occur, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. One indication that two nucleic acid sequences are substantially identical is when the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

(e)(ii) The term "substantial identity" in the context of a peptide indicates that a peptide comprises a sequence with at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, or 79%, preferably 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%, more preferably at least 90%, 91%, 92%, 93%, or 94%, or even more preferably, 95%, 96%, 97%, 98% or 99%, sequence identity to the reference sequence over a specified comparison window. Preferably, optimal alignment is conducted using the homology alignment algorithm of Needleman and Wunsch, 1970. An indication that two peptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the second peptide. Thus, a peptide is substantially identical to a second peptide, for example, where the two peptides differ only by a conservative substitution.

For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

As noted above, another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. The phrase "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (e.g., total cellular) DNA or RNA. "Bind(s)

substantially" refers to complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target nucleic acid sequence.

"Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and Northern hybridizations are sequence dependent, and are different under different environmental parameters. Longer sequences hybridize specifically at higher temperatures. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl, 1984; T_m 81.5°C + 16.6 (log M) +0.41 (%GC) - 0.61 (% form) - 500/L; where M is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. T_m is reduced by about 1°C for each 1% of mismatching; thus, T_m, hybridization, and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with >90% identity are sought, the T_m can be decreased 10°C. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4°C lower than the thermal melting point (T_m); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10°C lower than the thermal melting point (T_m); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20°C lower than the thermal melting point (T_m). Using the equation, hybridization and wash compositions, and desired T, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a T of less than 45°C (aqueous solution) or 32°C (formamide solution), it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993). Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH.

An example of highly stringent wash conditions is 0.15 M NaCl at 72°C for about 15 minutes. An example of stringent wash conditions is a 0.2X SSC wash at 65°C for 15 minutes (see, Sambrook, *infra*, for a description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1X SSC at 45°C for 15 minutes. An example low stringency wash for a duplex of, e.g., more than 100 nucleotides, is 4-6X SSC at 40°C for 15 minutes. For short probes (e.g., about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less than about 1.5 M, more preferably about 0.01 to 1.0 M, Na ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is typically at least about 30°C and at least about 60°C for long robes (e.g., >50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2X (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the proteins that they encode are substantially identical. This occurs, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or Northern blot is 50% formamide, e.g., hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0. 1X SSC at 60 to 65°C. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1.0 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C.

The following are examples of sets of hybridization/wash conditions that may be used to clone orthologous nucleotide sequences that are substantially identical to reference nucleotide sequences of the present invention: a reference nucleotide sequence preferably hybridizes to the reference nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C, more desirably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1% SDS at 50°C, more desirably still in 7% sodium dodecyl sulfate (SDS), 0.5 M

$NaPO_4$, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C, preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C, more preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C.

By "variant" polypeptide is intended a polypeptide derived from the native protein by deletion (so-called truncation) or addition of one or more amino acids to the N-terminal and/or C-terminal end of the native protein; deletion or addition of one or more amino acids at one or more sites in the native protein; or substitution of one or more amino acids at one or more sites in the native protein. Such variants may results form, for example, genetic polymorphism or from human manipulation. Methods for such manipulations are generally known in the art.

Thus, the polypeptides of the invention may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants of the polypeptides can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide sequence alterations are well known in the art. See, for example, Kunkel (1985); Kunkel et al. (1987); U. S. Patent No. 4,873,192; Walker and Gaastra (1983), and the references cited therein. Guidance as to appropriate amino acid substitutions that do not affect biological activity of the protein of interest may be found in the model of Dayhoff et al. (1978). Conservative substitutions, such as exchanging one amino acid with another having similar properties, are preferred.

Thus, the genes and nucleotide sequences of the invention include both the naturally occurring sequences as well as mutant forms. Likewise, the polypeptides of the invention encompass both naturally occurring proteins as well as variations and modified forms thereof. Such variants will continue to possess the desired activity. The deletions, insertions, and substitutions of the polypeptide sequence encompassed herein are not expected to produce radical changes in the characteristics of the polypeptide. However, when it is difficult to predict the exact effect of the substitution, deletion, or insertion in advance of doing so, one skilled in the art will appreciate that the effect will be evaluated by routine screening assays.

Individual substitutions deletions or additions that alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 1%) in an encoded sequence are "conservatively modified variations," where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. The

following five groups each contain amino acids that are conservative substitutions for one another: Aliphatic: Glycine (G), Alanine (A), Valine (V), Leucine (L), Isoleucine (I); Aromatic: Phenylalanine (F), Tyrosine (Y), Tryptophan (W); Sulfur-containing: Methionine (M), Cysteine (C); Basic: Arginine (R), Lysine (K), Histidine (H); Acidic: Aspartic acid (D), Glutamic acid (E), Asparagine (N), Glutamine (Q). See also, Creighton (1984). In addition, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence are also "conservatively modified variations."

"Production tissue" refers to mature, harvestable tissue consisting of non-dividing, terminally-differentiated cells. It excludes young, growing tissue consisting of germline, meristematic, and not-fully-differentiated cells.

"Germline cells" refer to cells that are destined to be gametes and whose genetic material is heritable.

The word "plant" refers to any plant, particularly to seed plant, and "plant cell" is a structural and physiological unit of the plant, e.g., a cell which comprises a cell wall or a protoplast. The plant cell may be in form of an isolated single cell or a cultured cell, or as a part of higher organized unit such as, for example, a plant tissue, or a plant organ.

"Plant tissue" includes differentiated and undifferentiated tissues or plants, including but not limited to roots, stems, shoots, leaves, pollen, seeds, tumor tissue and various forms of cells and culture such as single cells, protoplast, embryos, and callus tissue. The plant tissue may be in plants or in organ, tissue or cell culture.

The term "altered plant trait" means any phenotypic or genotypic change in a transgenic plant relative to the wild-type or non-transgenic plant host.

The term "transformation" refers to the transfer of a nucleic acid fragment into the genome of a host cell, resulting in genetically stable inheritance. Host cells containing the transformed nucleic acid fragments are referred to as "transgenic" cells, and organisms comprising transgenic cells are referred to as "transgenic organisms". Examples of methods of transformation of plants and plant cells include *Agrobacterium*-mediated transformation (De Blaere et al., 1987) and particle bombardment technology (Klein et al., 1987; U.S. Patent No. 4,945,050). Whole plants may be regenerated from transgenic cells by methods well known to the skilled artisan (see, for example, Fromm et al., 1990).

"Transformed," "transgenic," and "recombinant" refer to a host cell or organism such as a bacterium or a plant into which a heterologous nucleic acid molecule has been introduced.

The nucleic acid molecule can be stably integrated into the genome generally known in the art and are disclosed in Sambrook et al. (1989). See also Innis et al. (1995); and Gelfand (1995); and Innis and Gelfand (1999). Known methods of PCR include, but are not limited to, methods using paired primers, nested primers, single specific primers, degenerate primers, gene-specific primers, vector-specific primers, partially mismatched primers, and the like. For example, "transformed," "transformant," and "transgenic" plants or calli have been through the transformation process and contain a foreign gene integrated into their chromosome. The term "untransformed" refers to normal plants that have not been through the transformation process.

A "transgenic plant" is a plant having one or more plant cells that contain an expression vector.

"Transiently transformed" refers to cells in which transgenes and foreign DNA have been introduced (for example, by such methods as *Agrobacterium*-mediated transformation or biolistic bombardment), but not selected for stable maintenance.

"Stably transformed" refers to cells that have been selected and regenerated on a selection media following transformation.

"Transient expression" refers to transgene expression in cells, e.g., after transformation with recombinant virus or by such methods as *Agrobacterium*-mediated transformation, electroporation, or biolistic bombardment, but not selected for its stable maintenance.

"Genetically stable" and "heritable" refer to chromosomally-integrated genetic elements that are stably maintained in the plant and stably inherited by progeny through successive generations.

"Primary transformant" and "T0 generation" refer to transgenic plants that are of the same genetic generation as the tissue which was initially transformed (i.e., not having gone through meiosis and fertilization since transformation).

"Secondary transformants" and the "T1, T2, T3, etc. generations" refer to transgenic plants derived from primary transformants through one or more meiotic and fertilization cycles. They may be derived by self-fertilization of primary or secondary transformants or crosses of primary or secondary transformants with other transformed or untransformed plants.

"Significant increase" is an increase that is larger than the margin of error inherent in the measurement technique, preferably an increase by about 2-fold or greater.

"Significantly less" means that the decrease is larger than the margin of error inherent in the measurement technique, preferably a decrease by about 2-fold or greater.

I. The Nucleic Acid Molecules of the Invention and Polypeptide Encoded Thereby

This invention relates to isolated plant nucleic acid molecules, sequences and segments (fragments), the expression of which is increased in plants with increased protein content or levels, as well as the endogenous plant promoters for those expressed molecules, sequences or segments. Preferred sources for the nucleic acid molecules of the invention include, but are not limited to, corn (*Zea mays*), *Brassica* sp. (e.g., *B. napus*, *B. rapa*, *B. juncea*), particularly those *Brassica* species useful as sources of seed oil, alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), millet (e.g., pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*)), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium barbadense*, *Gossypium hirsutum*), sweet potato (*Ipomoea batatus*), cassava (*Manihot esculenta*), coffee (*Cofea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea americana*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, barley, vegetables, ornamentals, and conifers; duckweed (*Lemna*, see WO 00/07210, which includes members of the family *Lemnaceae*. There are known four genera and 34 species of duckweed as follows: genus *Lemna (L. aequinoctialis, L. disperma, L. ecuadoriensis, L. gibba, L. japonica, L. minor, L. miniscula, L. obscura, L. perpusilla, L. tenera, L.* trisulca, L. turionifera, L. valdiviana); *genus* Spirodela (S. intermedia, S. polyrrhiza, S.punctata*); genus* Woffia *(Wa. angusta, Wa. arrhiza, Wa. australina, Wa. borealis, Wa. brasiliensis, Wa. columbiana, Wa. elongata, Wa. globosa, Wa. microscopica, Wa. neglecta)* and genus *Wofiella (Wl. caudata, Wl. denticulata, Wl. gladiata, Wl. hyalina, Wl. lingulata, Wl. repunda, Wl. rotunda, and Wl. neotropica*). Any other genera or species of *Lemnaceae*, if they exist, are also aspects of the present invention. *Lemna gibba, Lemna minor*, and *Lemna miniscula* are preferred, with *Lemna* minor and *Lemna miniscula* being most preferred. *Lemna* species can be classified using the taxonomic scheme described by Landolt, Biosystematic Investigation on the Family of Duckweeds: The family of Lemnaceae - A Monograph Study. Geobatanischen Institut ETH, Stiftung Rubel, Zurich (1986)); vegetables including tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca*

sativa), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus* spp.), and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals include azalea (*Rhododendron* spp.), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa* spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum. Conifers that may be employed in practicing the present invention include, for example, pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*), Douglas-fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga canadensis*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as silver fir (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow-cedar (*Chamaecyparis nootkatensis*). Leguminous plants include beans and peas. Beans include guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, etc. Legumes include, but are not limited to, *Arachis*, e.g., peanuts, *Vicia*, e.g., crown vetch, hairy vetch, adzuki bean, mung bean, and chickpea, *Lupinus*, e.g., lupine, trifolium, *Phaseolus*, e.g., common bean and lima bean, *Pisum*, e.g., field bean, *Melilotus*, e.g., clover, *Medicago*, e.g., alfalfa, Lotus, e.g., trefoil, lens, e.g., lentil, and false indigo, *Acacia*, aneth, artichoke, arugula, blackberry, canola, cilantro, clementines, escarole, eucalyptus, fennel, grapefruit, honey dew, jicama, kiwifruit, lemon, lime, mushroom, nut, okra, orange, parsley, persimmon, plantain, pomegranate, poplar, radiata pine, radicchio, Southern pine, sweetgum, tangerine, triticale, vine, yams, apple, pear, quince, cherry, apricot, melon, hemp, buckwheat, grape, raspberry, chenopodium, blueberry, nectarine, peach, plum, strawberry, watermelon, eggplant, pepper, cauliflower, Brassica, e.g., broccoli, cabbage, brussels sprouts, onion, carrot, leek, beet, broad bean, celery, radish, pumpkin, endive, gourd, garlic, snapbean, spinach, squash, turnip, asparagus, and zucchini and ornamental plants include impatiens, Begonia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Agertum, Amaranthus, Antihirrhinum, Aquilegia, Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossos, and Zinnia.

Other vegetable sources (and databases to identify orthologs of the invention) for the nucleic acid sequences of the invention include those are shown in Table 1.

Table 1

FAMILY	LATIN NAME	COMMON NAME	MAP REFERENCES RESOURCES	LINKS
Cucurbitaceae	*Cucumis sativus*	Cucumber		http://www.cucurbit.org/
	Cucumis melo	Melon		http://genome.cornell.edu/cgc/
	Citrullus lanatus	Watermelon		
	Cucurbita pepo	Squash – summer		
	Cucurbita maxima	Squash - winter		
	Cucurbita moschata	Pumpkin /butternut		
Total				http://www.nal.usda.gov/pgdic/Map_proj/

Solanaceae	*Lycopersicon esculentum*	Tomato	• 15x BAC on variety Heinz 1706 order from Clemson Genome center (www.genome.clemson.edu) • 11.6x BAC of L. cheesmanii (orginates from J. Giovannoni) available from Clemson genome center (www.genome.clemson.edu) • EST collection from TIGR (www.tigr.org/tdb/lgi/index.html) • EST collection from Clemsom Genome Center (www.genome.clemson.edu) • TAG 99:254-271, 1999 (esculentum x pennelli) • TAG 89:1007-1013, 1994 (peruvianum) • Plant Cell Reports 12:293-297, 1993	genome.cornell.edu/solgenes http://ars-genome.cornell.edu/cgi-bin/WebAce/webace?db=solgenes http://genome.cornell.edu/tgc/ http://tgrc.ucdavis.edu/

	Capsicum annuum	Pepper		http://neptune.netimages.com/~chile/science.html
	Capsicum frutescens	Chile pepper		
	Solanum melongena	Eggplant		
	(Nicotiana tabacum)	(Tobacco)		
	(Solanum tuberosum)	(Potato)		
	(Petunia x hybrida hort. ex E. Vilm.)	(Petunia)	4x BAC of Petunia hybrida 7984 available from Clemson genome center (www.genome.clemson.edu)	
Total				http://www.nal.usda.gov/pgdic/Map_proj/
Brassicaceae	*Brassica oleracea* L. var. *italica*	Broccoli		http://res.agr.ca/ecorc/cwmt/crucifer/traits/index.htm http://geneous.cit.cornell.edu/cabbage/aboutcab.html

	Brassica oleracea L. var. *capitata*	Cabbage		
	Brassica rapa	Chinese Cabbage		
	Brassica oleracea L. var. *botrytis*	Cauliflower		
	Raphanus sativus var. *niger*	Daikon		
	(Brassica napus)	(Oilseed rape)		http://ars-genome.cornell.edu/cgi-bin/WebAce/webace?db=brassicadb
		Arabidopsis	12x and 6x BACs on Columbia strain available from Clemson genome center (www.genome.clemson.edu)	http://ars-genome.cornell.edu/cgi-bin/WebAce/webace?db=agr
Total				http://www.nal.usda.gov/pgdic/Map_proj/
Umbelliferae	*Daucus carota*	Carrot		
Compositae	*Lactuca sativa*	Lettuce		
	Helianthus annuus	(Sunflower)		
Total				

Chenopodiaceae	*Spinacia oleracea*	Spinach		
	(Beta vulgaris)	(Sugar Beet)		
Total				
Leguminosae	*Phaseolus vulgaris*	Bean	4.3x BAC available from Clemson genome center (www.genome.clemson.edu)	http://ars-genome.cornell.edu/cgi-bin/WebAce/webace?db=beangenes
	Pisum sativum	Pea		
	(Glycine max)	(Soybean)	7.5x and 7.9x BACs available from Clemson genome center (www.genome.clemson.edu)	http://ars-genome.cornell.edu/cgi-bin/WebAce/webace?db=soybase
Total			http://www.nal.usda.gov/pgdic/Map_proj/	
Gramineae	*Zea mays*	Sweet Corn	Novartis BACs for Mo17 and B73 have been donated to Clemson Genome Center (www.genome.clemson.edu)	
	(Zea mays)	(Field Corn)		http://www.agron.missouri.edu/mnl/

Total			http://www.nal.usda.gov/pgdic/Map_proj/	
Liliaceae	*Allium cepa*	Onion		
		Leek		
		(Garlic)		
		(Asparagus)		
Total			http://www.nal.usda.gov/pgdic/Map_proj/	

Preferred forage and turf grass nucleic acid sources for use in the methods of the invention include alfalfa, orchard grass, tall fescue, perennial ryegrass, creeping bent grass, and redtop. Preferably, the nucleic acid sources are crop plants and in particular cereals (for example, corn, alfalfa, sunflower, rice, *Brassica*, canola, soybean, barley, soybean, sugarbeet, cotton, safflower, peanut, sorghum, wheat, millet, tobacco, etc.), and even more preferably corn and soybean.

According to one embodiment, the present invention is directed to a nucleic acid molecule comprising a nucleotide sequence isolated from any plant which encodes a polypeptide having at least 70% amino acid sequence identity to a polypeptide comprising SEQ ID NOs. 1-36 or a promoter for said nucleotide sequence. Thus, based on the nucleic acid sequences encoding the polypeptide of the present invention, orthologs of those sequences may be identified or isolated from the genome of any desired organism, preferably from another plant, according to well known techniques based on their sequence similarity to the coding sequences, e.g., hybridization, PCR or computer generated sequence comparisons. For example, all or a portion of a particular plant sequence is used as a probe that selectively hybridizes to other gene sequences present in a population of cloned genomic DNA fragments or cDNA fragments (i.e., genomic or cDNA libraries) from a chosen source organism. Further, suitable genomic and cDNA libraries may be prepared from any cell or tissue of an organism. Such techniques include hybridization screening of plated DNA libraries (either plaques or colonies; see, e.g., Sambrook et al., 1989) and amplification by PCR using oligonucleotide primers preferably corresponding to sequence domains conserved among

related polypeptide or subsequences of the nucleotide sequences provided herein (see, e.g., Innis et al., 1990). These methods are particularly well suited to the isolation of gene sequences from organisms closely related to the organism from which the probe sequence is derived. The application of these methods using the coding sequences as probes is well suited for the isolation of gene sequences from any source organism, preferably other plant species. In a PCR approach, oligonucleotide primers can be designed for use in PCR reactions to amplify corresponding DNA sequences from cDNA or genomic DNA extracted from any plant of interest. Methods for designing PCR primers and PCR cloning are generally known in the art as discussed hereinabove.

In hybridization techniques, all or part of a known nucleotide sequence is used as a probe that selectively hybridizes to other corresponding nucleotide sequences present in a population of cloned genomic DNA fragments or cDNA fragments (i.e., genomic or cDNA libraries) from a chosen organism. The hybridization probes may be genomic DNA fragments, cDNA fragments, RNA fragments, or other oligonucleotides, and may be labeled with a detectable group such as ^{32}P, or any other detectable marker. Thus, for example, probes for hybridization can be made by labeling synthetic oligonucleotides based on the sequence of the invention. Methods for preparation of probes for hybridization and for construction of cDNA and genomic libraries are generally known in the art and are disclosed in Sambrook et al. (1989). In general, sequences that hybridize to the sequences disclosed herein will have at least 40% to 50%, about 60% to 70% and even about 80% 85%, 90%, 95% to 98% or more identity with the disclosed sequences. That is, the sequence similarity of sequences may range, sharing at least about 40% to 50%, about 60% to 70%, and even about 80%, 85%, 90%, 95% to 98% sequence similarity.

The nucleic acid molecules of the invention can also be identified by, for example, a search of known databases for genes encoding polypeptides having a specified amino acid sequence identity. Methods of alignment of sequences for comparison are well known in the art and are described hereinabove.

Eleven proteins and their orthologs, of the invention, and their sequences are listed in the Sequence Listing, and are further described. Globulin-1 s allele precursor and Globulin-2 precursor are embryo storage protein. Reference describing these two gene family include (1) Biochem Genet 1989 Apr;27(3-4):239-51 Characterization of embryo globulins encoded by the maize Glb genes. Kriz AL. and (2) Characterization of the maize Globulin-2 gene and analysis of two null alleles. Kriz AL, Wallace NH Biochem Genet 1991 Jun;29(5-6):241-54,

which are incorporated by reference. Oleosin is a proteins associated with seed oil body. It is also an ABA inducible protein, further described in (1) Frandsen GI, Mundy J, Tzen JT. Oil bodies and their associated proteins, oleosin and caleosin. Physiol Plant. 2001 Jul;112(3):301-307 and (2) Crowe AJ, Abenes M, Plant A, Moloney MM. The seed-specific transactivator, ABI3, induces oleosin gene expression. Plant Sci. 2000 Feb 21;151(2):171-181. which are incorporated by reference. The 17.2 KD heat shock protein is a stress induced protein, and are further described in Heat shock proteins, Martin E Feder and Gretchen E. Hofmann 1999 Molecular cahperones, and the stress responses : Evolutionary and ecological physiology. Annu Rev Physiol 61:243-282, which is incorporated by reference. Glucose and ribitol dehydrogenase homolog is an embryo-specifc protein, up-regulated during seed maturation, and is further described in Alexander R, Alamillo JM, Salamini F, Bartels D Planta 1994;192(4):519-25 A novel embryo-specific barley cDNA clone encodes a protein with homologies to bacterial glucose and ribitol dehydrogenase, which is incorporated by reference. ZMPK1 precursor is a putative receptor protein kinase related to stress response, and further described in Zhang R, Walker JC (1993) Structure and expression of the *S* locus-related genes of maize. Plant Mol Biol 21: 1171-1174, whichis incorporated herein by reference. Glutathione S-transferase is an enzyme for transferring glutathione to many substrates, including cytotoxic substances, and is further described in Marrs K.A. The function and regulation of glutathione S- transferases in plants, Annu. Rev. Plant Physiol. Plant. Mol. Biol. 1996 , Vol. 47: 127-158, which is incorporated herein by reference. Thioredoxin dependent peroxidase is an enzyme involved in Antioxidative Defence System, further described in RB Van Huystee, Some Molecular Aspects Of Plant Peroxidase Biosynthetic Studies, Annu. Rev. Plant Physiol. Plant. Mol. Biol. 1987 , Vol. 38: 205-219, which is incorporated herein by reference. RAB28 protein is an ABA induced gene, in late embryogenesis in repsponse to water stress, further described in Pla M, Gomez J, Goday A, Pagés M Regulation of the abscisic acid-responsive gene rab28 in maize viviparous mutants, Mol Gen Genet. 1991 Dec;230(3):394-400, which is incoproated herein by reference. Dehydrin dHN1 belongs to a group of proteins that are stress induced and involved in stress tolerance, further described in Zeevaart JAD, Creelman RA 1988 Metabolisms and physiology of abscisic acids. Annu Rev Plant Physiol Mol Biol 39:439-473, which is incoproated herein by reference. Hydroxymethylglutaryl-CoA reductase which is a key enzyme involved in catalyzing an early reaction unique to isoprenoid biosynthesis., further described in (1) Kato-Emori S, Higashi K, Hosoya K, Kobayashi T, Ezura H. Cloning and characterization of the gene encoding 3-

hydroxy-3-methylglutaryl coenzyme A reductase in melon (Cucumis melo L. reticulatus). Mol Genet Genomics. 2001 Mar;265(1):135-42, and (2) Caelles C, Ferrer A, Balcells L, Hegardt FG, Boronat A. Isolation and structural characterization of a cDNA encoding Arabidopsis thaliana 3-hydroxy-3-methylglutaryl coenzyme A reductase. Plant Mol Biol. 1989 Dec;13(6):627-38, which are incorporated herein by reference.

As described in the Examples, a proteomics approach was used to identify genes that were differentially expressed in high protein corn lines. Over 150 differentially expressed protein spots were identified and analyzed as described in the experimental conditions. Provded herein are genes, their proteins, as shown in the Sequence Listing, and their orthologs, applicable to the methods and compositons of the present invention.

The nature of the candidate genes and their potential roles in contributing to the high protein phenotype is presented, however, the inventor is by no means to be limited by any one proposed mechanism. Among the proteins positively annotated, two groups of proteins are outstanding and are believed to be intimately related to the corn high protein phenotype: one group represents the seed storage proteins, including globulins and oleosin. These major seed protein storage components are believed to directly contribute to the high protein phenotype. A second group of proteins can be roughly characterized as stress related proteins, such as the heat shock protein, dehydrin and a regulatory gene rab28 involved in ABA related stress response.

These two groups of proteins or genes are part of the same mechanism that contributes to the high protein yield. For example, ABI3, which is a Arabidopsis gene that involved in seed storage protein biosynthesis, is reported as a key player in temperature stress. One hypothesis for this relationship is that plants that are more stress resistant, such as more heat tolerant, will grow better, therefore have more grain yield including grain protein. Based on this hypothesis we postulate that other global regulators that have a significant impact on stress related response can be used to manipulate seed protein content, such as the ABI3 from Arabidopsis and its homolog in rice, and in particular the genes of the present invention.

Additional data presented herein support these relationships and uses of these protein and genes as to modulate high protein trait. Details are provided in th Examples section. Seed storage proteins directly contribute to the high protein phenotype. Antibodies developed against the two embryo specific globulin proteins, glb1 and 2 (see Sequence Listing) were used to deterimine that their protein levels in the high protein inbred lines are significantly higher than in the control line (1.5-2 fold). Gene expression patterns of selected genes were studied

in rice gene expression profiling experiments. All three genes studied, the heat shock 17 gene, the glucose dehydrogenase gene and dehydrin gene (see Sequence Listing) were up-regulated in the time cource of rice seed development, coinciding with the development phase during which seed maturation occurs, indicating they play an important role during grain filling. (see Table 2) We also took a genetic approach to study the segregation of some of the genes with the high protein phenotype. A population of hybrid corn lines were used, derived from high protein line Wil500 as the high protein source. Table 3 demonstrates the correlation between the high protein trait and the HS 17 gene expression level.

Both seed protein content and protein quality can be changed by using these genes. In one embodiment a transgenic approach to over express or down regulate these genes in the seeds can be employed to increase grain protein content. Meanwhile, as is evident from their amino acid sequences (see Sequence Listing), some of these genes and proteins are biased to a special amino acid profile, over expressing of these proteins in seed can change the seed protein property. In particular, nutrionally enhanced seed, more complete or elevated in one or more amino acids, can be obtained. For example, poultry, like swine, have a specific amino acid that, if deficient, will reduce the animal's performance on the feed. For poultry, the limiting amino acid is methionine, while for swine the limiting amino acid is lysine. Accordingly, the present invention can provide feed that contains, or can be formulated to contain, an increase in the amount of methionine (or lysine for swine) and general protein to keep the desired protein gross energy ratio in the diet.

The genes disclosed herein are useful as genetic markers in marker assisted breeding programs to select high protein lines during plant breeding. For example, antibodies to the proteins of the invention (for use in ELISA for example) or DNA markers that are linked to these genes, or the genes themselves, find use to predict genomic profile and thus trait outcome of siblings in breeding practices.

Furthermore, the genes and proteins find use in production of effective protein production factories. For example, down regulation of one or more of the proteins can provide a seed that is reduced in these proteins thus allowing increased cellular resources for expression of industrially or therapeutically important polypeptides. This can best be done by inducible regulation of the one or more genes. In one embodiment reduction of storage protein content is achieved by anti-sense, for example RNAi methods. A two component Gal4/C1 system can be used to provide an inducible system. Two components in separate inbred lines are inactive, but create hybrids in which gene mosulation is activated to create plants with

lower protein yield. This methd also finds use to create low protein lines that are less allergenic.

Further details for use of these genes and proteins and their orthologs are presented herein.

II. Expression Cassettes of the Invention

The present invention also encompasses expression cassettes, preferably in the form of a recombinant vectors comprising the nucleic acid sequences of the invention. In such vectors, the expression cassette comprises regulatory elements for expression of the nucleotide sequences in a host cell capable of expressing the nucleotide sequences. Such regulatory elements usually comprise promoter and termination signals and preferably also comprise elements allowing efficient translation of polypeptides encoded by the nucleic acid sequences of the present invention. For efficient initiation of translation, sequences adjacent to the initiating methionine may require modification. For example, they can be modified by the inclusion of sequences known to be effective in plants. Joshi (1987) has suggested an appropriate consensus for plants and Clontech suggests a further consensus translation initiator (1993/1994 catalog, page 210). These consensuses are suitable for use with the nucleotide sequences of this invention. The sequences are incorporated into constructions comprising the nucleotide sequences, up to and including the ATG (whilst leaving the second amino acid unmodified), or alternatively up to and including the GTC subsequent to the ATG (with the possibility of modifying the second amino acid of the transgene).

Vectors comprising the nucleic acid sequences are usually capable of replication in particular host cells, e.g., as extrachromosomal molecules, and are therefore used to amplify the nucleic acid sequences of this invention in the host cells. In a preferred embodiment, host cells for such vectors are plant cells.

A. Promoters and Enhancers

Expression of the nucleotide sequences in transgenic plants is driven by promoters shown to be functional in plants. The choice of promoter will vary depending on the temporal and spatial requirements for expression, and also depending on the target species. In many cases, expression in multiple tissues is desirable. Although many promoters from dicotyledons have been shown to be operational in monocotyledons and *vice versa*, ideally dicotyledonous promoters are selected for expression in dicotyledons, and monocotyledonous promoters for expression in monocotyledons. However, there is no restriction to the provenance of selected

promoters; it is sufficient that they are operational in driving the expression of the nucleotide sequences in the desired cell.

These promoters include, but are not limited to, constitutive, inducible, temporally regulated, developmentally regulated, chemically regulated, stress-responsive, tissue-preferred and tissue-specific promoters. Promoter sequences are known to be strong or weak. A strong promoter provides for a high level of gene expression, whereas a weak promoter provides for a very low level of gene expression. An inducible promoter is a promoter that provides for the turning on and off of gene expression in response to an exogenously added agent, or to an environmental or developmental stimulus. A bacterial promoter such as the P_{tac} promoter can be induced to varying levels of gene expression depending on the level of isothiopropylgalactoside added to the transformed bacterial cells. An isolated promoter sequence that is a strong promoter for heterologous nucleic acid is advantageous because it provides for a sufficient level of gene expression to allow for easy detection and selection of transformed cells and provides for a high level of gene expression when desired.

Preferred promoters that are expressed constitutively include promoters from genes encoding actin or ubiquitin and the CaMV 35S and 19S promoters. The nucleotide sequences of this invention can also be expressed under the regulation of promoters that are chemically regulated. This enables the nucleic acid sequence or encoded polypeptide to be synthesized only when the crop plants are treated with the inducing chemicals. Preferred technology for chemical induction of gene expression is detailed in the published application EP 0 332 104 (to Ciba-Geigy) and U.S. Patent 5,614,395. A preferred promoter for chemical induction is the tobacco PR-1a promoter.

Tissue-specific or tissue-preferential promoters useful in the present invention. Also useful are promoters which confer seed-specific expression, such as those disclosed by Schernthaner et al. (1988); anther (tapetal) specific promoter B6 (Huffman et al.); and pistil-specific promoters such as a modified S13 promoter (Dzelkalns et al., 1993).

Preferred tissue specific expression patterns include green tissue-specific, root-specific, stem-specific, and flower-specific. Promoters suitable for expression in green tissue include many which regulate genes involved in photosynthesis and many of these have been cloned from both monocotyledons and dicotyledons. A preferred promoter is the maize PEPC promoter from the phosphoenol carboxylase gene (Hudspeth & Grula, 1989). A preferred promoter for root-specific expression is that described by de Framond (1991; EP 0 452 269 to

Ciba-Geigy). A preferred stem specific promoter is that described in U.S. Patent No. 5,625,136 (to Ciba-Geigy) and which drives expression of the maize *trpA* gene. .

Other promoters which direct specific or enhanced expression in certain plant tissues will be known to those of skill in the art in light of the present disclosure. These include, for example, the *rbcS* promoter, specific for green tissue; the *ocs, nos*, and *mas* promoters which have higher activity in roots or wounded leaf tissue; a truncated (-90 to +8) 35S promoter which directs enhanced expression in roots, an tubulin gene that directs expression in roots and promoters derived from zein storage protein genes which direct expression in endosperm. It is particularly contemplated that one may advantageously use the 16 bp *ocs* enhancer element from the octopine synthase (*ocs*) gene (Bonchez et al., 1989), especially when present in multiple copies, to achieve enhanced expression in roots.

Preferred plant promoters include, but are not limited to, a promoter such as the CaMV 35S promoter, an enhanced 35S promoter or others such as CaMV 19S, *nos*, *Adh1*, sucrose synthase, ∀-tubulin, ubiquitin, actin, *cab*, PEPCase or those associated with the R gene complex. Further suitable promoters may include the U2 and U5 snRNA promoters from maize, the promoter from alcohol dehydrogenase, the Z4 promoter from a gene encoding the Z4 22 kD zein protein, the Z10 promoter from a gene encoding a 10 kD zein protein, a Z27 promoter from a gene encoding a 27 kD zein protein, the A20 promoter from the gene encoding a 19 kD -zein protein, inducible promoters, such as the light inducible promoter derived from the pea rbcS gene and the actin promoter from rice; seed specific promoters, such as the phaseolin promoter from beans, may also be used. Other promoters useful in the practice of the invention are known to those of skill in the art.

Examples of tissue specific promoters which have been described include the lectin (Vodkin, 1983; Lindstrom et al., 1990,) corn alcohol dehydrogenase 1 (Vogel et al., 1992; Dennis et al., 1984), corn light harvesting complex (Simpson, 1985; Bansal et al., 1992), corn heat shock protein (Odell et al., 1985; Rochester et al., 1986), pea small subunit RuBP carboxylase (Poulsen et al., 1986; Cashmore et al., 1983), Ti plasmid mannopine synthase (Langridge et al., 1989), Ti plasmid nopaline synthase (Langridge et al., 1989), petunia chalcone isomerase (vanTunen et al., 1988), bean glycine rich protein 1 (Keller et al., 1989), truncated CaMV 35s (Odell et al., 1985), potato patatin (Wenzler et al., 1989), root cell (Yamamoto et al., 1990), maize zein (Reina et al., 1990; Kriz et al., 1987; Wandelt et al., 1989; Langridge et al., 1983; Reina et al., 1990), globulin-1 (Belanger et al., 1991), α-tubulin,

cab (Sullivan et al., 1989), PEPCase (Hudspeth & Grula, 1989), R gene complex-associated promoters (Chandler et al., 1989), and chalcone synthase promoters (Franken et al., 1991).

Inducible promoters that have been described include the ABA- and turgor-inducible promoters, the promoter of the auxin-binding protein gene (Schwob et al., 1993), the UDP glucose flavonoid glycosyl-transferase gene promoter (Ralston et al., 1988), the MPI proteinase inhibitor promoter (Cordero et al., 1994), and the glyceraldehyde-3-phosphate dehydrogenase gene promoter (Kohler et al., 1995; Quigley et al., 1989; Martinez et al., 1989).

Several tissue-specific regulated genes and/or promoters have been reported in plants. These include genes encoding the seed storage proteins (such as napin, cruciferin, beta-conglycinin, and phaseolin) zein or oil body proteins (such as oleosin), or genes involved in fatty acid biosynthesis (including acyl carrier protein, stearoyl-ACP desaturase. and fatty acid desaturases (fad 2-1)), and other genes expressed during embryo development (such as Bce4, see, for example. EP 255378 and Kridl et al., 1991). Particularly useful for seed-specific expression is the pea vicilin promoter (Czako et al., 1992). (See also U.S. Pat. No. 5,625,136, herein incorporated by reference.) Other useful promoters for expression in mature leaves are those that are switched on at the onset of senescence, such as the SAG promoter from Arabidopsis (Gan et al., 1995, 270 (5244), 1986-8).

A class of fruit-specific promoters expressed at or during antithesis through fruit development, at least until the beginning of ripening, is discussed in U.S. 4,943,674, the disclosure of which is hereby incorporated by reference. cDNA clones that are preferentially expressed in cotton fiber have been isolated (John et al., 1992). cDNA clones from tomato displaying differential expression during fruit development have been isolated and characterized (Mansson et al., 1985, Slater et al., 1985). The promoter for polygalacturonase gene is active in fruit ripening. The polygalacturonase gene is described in U.S. Patent No. 4,535,060, U.S. Patent No. 4,769,061, U.S. Patent No. 4,801,590, and U.S. Patent No. 5,107,065, which disclosures are incorporated herein by reference.

Other examples of tissue-specific promoters include those that direct expression in leaf cells following damage to the leaf (for example, from chewing insects), in tubers (for example, patatin gene promoter), and in fiber cells (an example of a developmentally-regulated fiber cell protein is E6 (John et al., 1992). The E6 gene is most active in fiber, although low levels of transcripts are found in leaf, ovule and flower.

The tissue-specificity of some "tissue-specific" promoters may not be absolute and may be tested by one skilled in the art using the diphtheria toxin sequence. One can also achieve

tissue-specific expression with "leaky" expression by a combination of different tissue-specific promoters (Beals et al., 1997). Other tissue-specific promoters can be isolated by one skilled in the art (see U.S. 5,589,379). Several inducible promoters ("gene switches") have been reported. Many are described in the review by Gatz (1996 and 1997). These include tetracycline repressor system, *Lac* repressor system, copper-inducible systems, salicylate-inducible systems (such as the PR1a system), glucocorticoid- (Aoyama, 1997) and ecdysome-inducible systems. Also included are the benzene sulphonamide- (U.S. Patent No. 5,364,780) and alcohol- (WO 97/06269 and WO 97/06268) inducible systems and glutathione S-transferase promoters. Other studies have focused on genes inducibly regulated in response to environmental stress or stimuli such as increased salinity, drought, pathogen and wounding. (Graham et al., 1985; Graham et al., 1985, Smith et al., 1986). Accumulation of metallocarboxypeptidase-inhibitor protein has been reported in leaves of wounded potato plants (Graham et al., 1981). Other plant genes have been reported to be induced methyl jasmonate, elicitors, heat-shock, anaerobic stress, or herbicide safeners.

Frequently it is desirable to have continuous or inducible expression of a DNA sequence throughout the cells of an organism in a tissue-independent manner. For example, increased resistance of a plant to infection by soil- and air borne pathogens might be accomplished by genetic manipulation of the plant's genome to comprise a continuous promoter operably linked to a heterologous or homologous pathogen-resistance gene such that pathogen-resistance proteins are continuously expressed throughout the plant's tissues.

Alternatively, it might be desirable to inhibit expression of a native DNA sequence within a plant's tissues to achieve a desired phenotype. In this case, such inhibition might be accomplished with transformation of the plant to comprise a constitutive, tissue-independent promoter operably linked to an antisense nucleotide sequence, such that constitutive expression of the antisense sequence produces an RNA transcript that interferes with translation of the mRNA of the native DNA sequence.

Other elements include those that can be regulated by endogenous or exogenous agents, e.g., by DNA binding proteins such as zinc finger proteins, including naturally occurring zinc finger proteins or chimeric zinc finger proteins (see, e.g., U.S. Patent No. 5,789,538, WO 99/48909; WO 99/45132; WO 98/53060; WO 98/53057; WO 98/53058; WO 00/23464; WO 95/19431; and WO 98/54311) or myb-like transcription factors. For example, a chimeric zinc finger protein may include amino acid sequences which bind to a specific DNA sequence (the

zinc finger) and amino acid sequences that activate (e.g., GAL 4 sequences) or repress the transcription of the sequences linked to the specific DNA sequence.

B. 5' and 3' Sequences

In addition to promoters, a variety of 3 transcriptional terminators are also available for use in the present invention. Transcriptional terminators are responsible for the termination of transcription and correct mRNA polyadenylation. The 3N nontranslated regulatory DNA sequence preferably includes from about 50 to about 1,000, more preferably about 100 to about 1,000, nucleotide base pairs and contains plant transcriptional and translational termination sequences. Appropriate transcriptional terminators and those which are known to function in plants include the CaMV 35S terminator, the tml terminator, the nopaline synthase terminator, the pea rbcS E9 terminator, the terminator for the T7 transcript from the octopine synthase gene of *Agrobacterium tumefaciens*, and the 3N end of the protease inhibitor I or II genes from potato or tomato, although other 3N elements known to those of skill in the art can also be employed.

The 5N regulatory region of the expression cassette may also include other enhancing sequences. Numerous sequences have been found to enhance gene expression in transgenic plants. These include sequences which have been shown to enhance expression such as intron sequences (e.g., from *Adh1, bronze1* or the sucrose synthase intron) and viral leader sequences (e.g., from TMV, MCMV and AMV). For example, a number of non-translated leader sequences derived from viruses are known to enhance expression. Specifically, leader sequences from Tobacco Mosaic Virus (TMV), Maize Chlorotic Mottle Virus (MCMV), and Alfalfa Mosaic Virus (AMV) have been shown to be effective in enhancing expression (e.g., Gallie et al., 1987; Skuzeski et al., 1990). Other leaders known in the art include but are not limited to: Picornavirus leaders, for example, EMCV leader (Encephalomyocarditis 5 noncoding region) (Elroy-Stein et al., 1989); Potyvirus leaders, for example, TEV leader (Tobacco Etch Virus) (Allison et al., 1986); MDMV leader (Maize Dwarf Mosaic Virus); Human immunoglobulin heavy-chain binding protein (BiP) leader, (Macejak et al., 1991); Untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4), (Jobling et al., 1987; Tobacco mosaic virus leader (TMV), (Gallie et al., 1989; and Maize Chlorotic Mottle Virus leader (MCMV) (Lommel et al., 1991. See also, Della-Cioppa et al., 1987.

C. Targeting Sequences

It may be preferable to target expression of the nucleotide sequences of the present invention to different cellular localizations in the plant. In some cases, localization in the cytosol may be desirable, whereas in other cases, localization in some subcellular organelle, e.g., the nucleus, may be preferred. Subcellular localization of transgene encoded enzymes is undertaken using techniques well known in the art. Typically, the DNA encoding the target peptide from a known organelle-targeted gene product is manipulated and fused upstream of the nucleotide sequence. Many such target sequences are known for the chloroplast and their functioning in heterologous constructions has been shown. The expression of the nucleotide sequences of the present invention is also targeted to the endoplasmic reticulum or to the vacuoles of the host cells. Techniques to achieve this are well-known in the art.

D. Marker Genes

In order to improve the ability to identify transformants, one may desire to employ a selectable or screenable marker gene as, or in addition to, the preselected nucleic acid sequence or segment. "Marker genes" are genes that impart a distinct phenotype to cells expressing the marker gene and thus allow such transformed cells to be distinguished from cells that do not have the marker. Such genes may encode either a selectable or screenable marker, depending on whether the marker confers a trait which one can 'select' for by chemical means, i.e., through the use of a selective agent (e.g., a herbicide, antibiotic, or the like), or whether it is simply a trait that one can identify through observation or testing, i.e., by 'screening' (e.g., the R-locus trait). Of course, many examples of suitable marker genes are known to the art and can be employed in the practice of the invention.

Included within the terms selectable or screenable marker genes are also genes which encode a "secretable marker" whose secretion can be detected as a means of identifying or selecting for transformed cells. Examples include markers which encode a secretable antigen that can be identified by antibody interaction, or even secretable enzymes which can be detected by their catalytic activity. Secretable proteins fall into a number of classes, including small, diffusible proteins detectable, e.g., by ELISA; small active enzymes detectable in extracellular solution (e.g., ∀-amylase, ∃-lactamase, phosphinothricin acetyltransferase); and proteins that are inserted or trapped in the cell wall (e.g., proteins that include a leader sequence such as that found in the expression unit of extensin or tobacco PR-S).

With regard to selectable secretable markers, the use of a gene that encodes a polypeptide that becomes sequestered in the cell wall, and which polypeptide includes a

unique epitope is considered to be particularly advantageous. Such a secreted antigen marker would ideally employ an epitope sequence that would provide low background in plant tissue, a promoter-leader sequence that would impart efficient expression and targeting across the plasma membrane, and would produce protein that is bound in the cell wall and yet accessible to antibodies. A normally secreted wall protein modified to include a unique epitope would satisfy all such requirements.

Numerous other possible selectable and/or screenable marker genes will be apparent to those of skill in the art in addition to the one set forth herein below. Therefore, it will be understood that the following discussion is exemplary rather than exhaustive. In light of the techniques disclosed herein and the general recombinant techniques which are known in the art, the present invention renders possible the introduction of any gene, including marker genes, into a recipient cell to generate a transformed plant cell, e.g., a monocot cell.

Possible selectable markers for use in connection with the present invention include, but are not limited to, a *neo* gene, which codes for kanamycin resistance and can be selected for using kanamycin, G418, a gene encoding resistance to bleomycin, and the like; a *bar* gene which codes for bialaphos resistance; a gene which encodes an altered EPSP synthase protein thus conferring glyphosate resistance; a nitrilase gene such as *bxn* from *Klebsiella ozaenae* which confers resistance to bromoxynil; a mutant acetolactate synthase gene (ALS) which confers resistance to imidazolinone, sulfonylurea or other ALS-inhibiting chemicals (European Patent Application 154,204, 1985); a methotrexate-resistant DHFR gene; a dalapon dehalogenase gene that confers resistance to the herbicide dalapon; or a mutated anthranilate synthase gene that confers resistance to 5-methyl tryptophan. Where a mutant EPSP synthase gene is employed, additional benefit may be realized through the incorporation of a suitable chloroplast transit peptide, CTP (European Patent Application 0 218 571, 1987).

An illustrative embodiment of a selectable marker gene capable of being used in systems to select transformants is the genes that encode the enzyme phosphinothricin acetyltransferase, such as the *bar* gene from *Streptomyces hygroscopicus* or the *pat* gene from *Streptomyces viridochromogenes* (U.S. Patent No. 5,550,318). The enzyme phosphinothricin acetyltransferase (PAT) inactivates the active ingredient in the herbicide bialaphos, phosphinothricin (PPT). PPT inhibits glutamine synthetase, causing rapid accumulation of ammonia and cell death. The success in using this selective system in conjunction with monocots was particularly surprising because of the major difficulties which have been reported in transformation of cereals.

Screenable markers that may be employed include, but are not limited to, a ∃-glucuronidase or *uid*A gene (GUS) which encodes an enzyme for which various chromogenic substrates are known; an R-locus gene, which encodes a product that regulates the production of anthocyanin pigments (red color) in plant tissues; a -lactamase gene, which encodes an enzyme for which various chromogenic substrates are known (e.g., PADAC, a chromogenic cephalosporin); a *xyl*E gene which encodes a catechol dioxygenase that can convert chromogenic catechols; an -amylase gene; a tyrosinase gene which encodes an enzyme capable of oxidizing tyrosine to DOPA and dopaquinone which in turn condenses to form the easily detectable compound melanin; a -galactosidase gene, which encodes an enzyme for which there are chromogenic substrates; a luciferase (*lux*) gene, which allows for bioluminescence detection; or an aequorin gene, which may be employed in calcium-sensitive bioluminescence detection, or a green fluorescent protein.

Genes from the maize R gene complex are contemplated to be particularly useful as screenable markers. The R gene complex in maize encodes a protein that acts to regulate the production of anthocyanin pigments in most seed and plant tissue. Maize strains can have one, or as many as four, R alleles which combine to regulate pigmentation in a developmental and tissue specific manner. A gene from the R gene complex was applied to maize transformation, because the expression of this gene in transformed cells does not harm the cells. Thus, an R gene introduced into such cells will cause the expression of a red pigment and, if stably incorporated, can be visually scored as a red sector. If a maize line carries dominant alleles for genes encoding the enzymatic intermediates in the anthocyanin biosynthetic pathway (C2, A1, A2, Bz1 and Bz2), but carries a recessive allele at the R locus, transformation of any cell from that line with R will result in red pigment formation. Exemplary lines include Wisconsin 22 which contains the rg-Stadler allele and TR112, a K55 derivative which is r-g, b, Pl. Alternatively any genotype of maize can be utilized if the C1 and R alleles are introduced together.

A further screenable marker contemplated for use in the present invention is firefly luciferase, encoded by the *lux* gene. The presence of the *lux* gene in transformed cells may be detected using, for example, X-ray film, scintillation counting, fluorescent spectrophotometry, low-light video cameras, photon counting cameras or multiwell luminometry. It is also envisioned that this system may be developed for populational screening for bioluminescence, such as on tissue culture plates, or even for whole plant screening.

E. Other Sequences

A vector of the invention can also further comprise plasmid DNA. Plasmid vectors include additional DNA sequences that provide for easy selection, amplification, and transformation of the expression cassette in prokaryotic and eukaryotic cells, e.g., pUC-derived vectors such as pUC8, pUC9, pUC18, pUC19, pUC23, pUC119, and pUC120, pSK-derived vectors, pGEM-derived vectors, pSP-derived vectors, or pBS-derived vectors. The additional DNA sequences include origins of replication to provide for autonomous replication of the vector, additional selectable marker genes, preferably encoding antibiotic or herbicide resistance, unique multiple cloning sites providing for multiple sites to insert DNA sequences or genes encoded in the expression cassette, and sequences that enhance transformation of prokaryotic and eukaryotic cells.

Another vector that is useful for expression in both plant and prokaryotic cells is the binary Ti plasmid (as disclosed in Schilperoort et al., U.S. Patent No. 4,940,838) as exemplified by vector pGA582. This binary Ti plasmid vector has been previously characterized by An, cited *supra*. This binary Ti vector can be replicated in prokaryotic bacteria such as *E. coli* and *Agrobacterium*. The *Agrobacterium* plasmid vectors can be used to transfer the expression cassette to dicot plant cells, and under certain conditions to monocot cells, such as rice cells. The binary Ti vectors preferably include the nopaline T DNA right and left borders to provide for efficient plant cell transformation, a selectable marker gene, unique multiple cloning sites in the T border regions, the *col*E1 replication of origin and a wide host range replicon. The binary Ti vectors carrying an expression cassette of the invention can be used to transform both prokaryotic and eukaryotic cells, but is preferably used to transform dicot plant cells.

Virtually any DNA may be used for delivery to recipient cells to ultimately produce fertile transgenic plants in accordance with the present invention. For example, DNA segments in the form of vectors and plasmids, or linear DNA fragments, in some instance containing only the DNA element to be expressed in the plant, and the like, may be employed.

Vectors, plasmids, cosmids, YACs (yeast artificial chromosomes) and DNA segments for use in transforming such cells will, of course, generally comprise the cDNA, gene or genes which one desires to introduce into the cells. These DNA constructs can further include structures such as promoters, enhancers, polylinkers, or even regulatory genes as desired. The DNA segment or gene chosen for cellular introduction will often encode a protein which will be expressed in the resultant recombinant cells, such as will result in a screenable or selectable trait and/or which will impart an improved phenotype to the regenerated plant. However, this

may not always be the case, and the present invention also encompasses transgenic plants incorporating non-expressed transgenes.

III. Transformation

The expression cassettes of the present invention can be introduced into a host cell, e.g., a plant cell, in a number of art-recognized ways. Those skilled in the art will appreciate that the choice of method might depend on the type of cell, e.g., monocotyledonous or dicotyledonous, targeted for transformation. Vectors which may be used to transform plant tissue with the expression cassettes of the present invention include both *Agrobacterium* vectors and ballistic vectors, as well as vectors suitable for DNA-mediated transformation, e.g., direct uptake or via electroporation. However, cells other than plant cells may be transformed with the expression cassettes of the invention.

Suitable methods of transforming plant cells include, but are not limited to, microinjection (Crossway et al., 1986), direct DNA transfer to plant cells by PEG precipitation; liposomes; electroporation (Riggs et al., 1986, *Agrobacterium*-mediated transformation (Hinchee et al., 1988), direct gene transfer (Paszkowski et al., 1984), and ballistic particle acceleration using devices available from Agracetus, Inc., Madison, Wis. and BioRad, Hercules, Calif. (see, for example, Sanford et al., U.S. Pat. No. 4,945,050; and McCabe et al., 1988). Also see, Weissinger et al., 1988; Sanford et al., 1987 (onion); Christou et al., 1988 (soybean); McCabe et al., 1988 (soybean); Datta et al., 1990 (rice); Klein et al., 1988 (maize); Klein et al., 1988 (maize); Klein et al., 1988 (maize); Fromm et al., 1990 (maize); and Gordon-Kamm et al., 1990 (maize); Svab et al., 1990 (tobacco chloroplast); Koziel et al., 1993 (maize); Shimamoto et al., 1989 (rice); Christou et al., 1991 (rice); European Patent Application EP 0 332 581 (orchardgrass and other Pooideae); Vasil et al., 1993 (wheat); Weeks et al., 1993 (wheat).

In one embodiment, a nucleotide sequence of the present invention is directly transformed into the plastid genome. Plastid transformation technology is extensively described in U.S. Patent Nos. 5,451,513, 5,545,817, and 5,545,818, in PCT application no. WO 95/16783, and in McBride et al., 1994. The basic technique for chloroplast transformation involves introducing regions of cloned plastid DNA flanking a selectable marker together with the gene of interest into a suitable target tissue, e.g., using biolistics or protoplast transformation (e.g., calcium chloride or PEG mediated transformation). The 1 to 1.5 kb flanking regions, termed targeting sequences, facilitate orthologous recombination with the

plastid genome and thus allow the replacement or modification of specific regions of the plastome. Initially, point mutations in the chloroplast 16S rRNA and rps12 genes conferring resistance to spectinomycin and/or streptomycin are utilized as selectable markers for transformation (Svab et al., 1990; Staub et al., 1992). This resulted in stable homoplasmic transformants at a frequency of approximately one per 100 bombardments of target leaves. The presence of cloning sites between these markers allowed creation of a plastid targeting vector for introduction of foreign genes (Staub et al., 1993). Substantial increases in transformation frequency are obtained by replacement of the recessive rRNA or r-protein antibiotic resistance genes with a dominant selectable marker, the bacterial *aadA* gene encoding the spectinomycin-detoxifying enzyme aminoglycoside-3N-adenyltransferase (Svab et al., 1993). Other selectable markers useful for plastid transformation are known in the art and encompassed within the scope of the invention. Typically, approximately 15-20 cell division cycles following transformation are required to reach a homoplastidic state. Plastid expression, in which genes are inserted by orthologous recombination into all of the several thousand copies of the circular plastid genome present in each plant cell, takes advantage of the enormous copy number advantage over nuclear-expressed genes to permit expression levels that can readily exceed 10% of the total soluble plant protein. In a preferred embodiment, a nucleotide sequence of the present invention is inserted into a plastid targeting vector and transformed into the plastid genome of a desired plant host. Plants homoplastic for plastid genomes containing a nucleotide sequence of the present invention are obtained, and are preferentially capable of high expression of the nucleotide sequence.

Agrobacterium tumefaciens cells containing a vector comprising an expression cassette of the present invention, wherein the vector comprises a Ti plasmid, are useful in methods of making transformed plants. Plant cells are infected with an *Agrobacterium tumefaciens* as described above to produce a transformed plant cell, and then a plant is regenerated from the transformed plant cell. Numerous *Agrobacterium* vector systems useful in carrying out the present invention are known. For example, U.S. Pat. No. 4,459,355 discloses a method for transforming susceptible plants, including dicots, with an *Agrobacterium* strain containing the Ti plasmid. The transformation of woody plants with an *Agrobacterium* vector is disclosed in U.S. Patent No. 4,795,855. Further, U.S. Patent No. 4,940,838 to Schilperoort et al. discloses a binary *Agrobacterium* vector (i.e., one in which the *Agrobacterium* contains one plasmid having the vir region of a Ti plasmid but no T region, and a second plasmid having a T region but no vir region) useful in carrying out the present invention.

It is particularly preferred to use the binary type vectors of Ti and Ri plasmids of *Agrobacterium spp.* Ti-derived vectors transform a wide variety of higher plants, including monocotyledonous and dicotyledonous plants, such as soybean, cotton, rape, tobacco, and rice (Pacciotti et al., 1985: Byrne et al., 1987; Sukhapinda et al., 1987; Lorz et al., 1985; Potrykus, 1985; Park et al., 1985: Hiei et al., 1994. The use of T-DNA to transform plant cells has received extensive study and is amply described (EP 120516; Hoekema, 1985; Knauf, et al., 1983; and An. et al., 1985. For introduction into plants, the nucleotide sequences of the invention can be inserted into binary vectors as described in the examples.

Transformation of plants can be undertaken with a single DNA molecule or multiple DNA molecules (i.e., co-transformation), and both these techniques are suitable for use with the expression cassettes of the present invention. Numerous transformation vectors are available for plant transformation, and the expression cassettes of this invention can be used in conjunction with any such vectors. The selection of vector will depend upon the preferred transformation technique and the target species for transformation.

Preferred plant cells for transformation include, but are not limited to, cells from plant such as corn (*Zea mays*), *Brassica* sp. (e.g., *B. napus*, *B. rapa*, *B. juncea*), particularly those *Brassica* species useful as sources of seed oil, alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), millet (e.g., pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*)), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium barbadense*, *Gossypium hirsutum*), sweet potato (*Ipomoea batatus*), cassava (*Manihot esculenta*), coffee (*Cofea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea americana*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, barley, vegetables, ornamentals, and conifers; duckweed (*Lemna*, see WO 00/07210, which includes members of the family *Lemnaceae*. There are known four genera and 34 species of duckweed as follows: genus *Lemna (L. aequinoctialis, L. disperma, L. ecuadoriensis, L. gibba, L. japonica, L. minor, L. miniscula, L. obscura, L. perpusilla, L. tenera, L.* trisulca, L. turionifera, L. valdiviana); genus *Spirodela (S. intermedia, S. polyrrhiza,*

S.punctata); genus *Woffia (Wa. angusta, Wa. arrhiza, Wa. australina, Wa. borealis, Wa. brasiliensis, Wa. columbiana, Wa. elongata, Wa. globosa, Wa. microscopica, Wa. neglecta)* and genus *Wofiella (Wl. caudata, Wl. denticulata, Wl. gladiata, Wl. hyalina, Wl. lingulata, Wl. repunda, Wl. rotunda, and Wl. neotropica)*. Any other genera or species of *Lemnaceae*, if they exist, are also aspects of the present invention. *Lemna gibba, Lemna minor*, and *Lemna miniscula* are preferred, with *Lemna* minor and *Lemna miniscula* being most preferred. *Lemna* species can be classified using the taxonomic scheme described by Landolt, Biosystematic Investigation on the Family of Duckweeds: The family of Lemnaceae - A Monograph Study. Geobatanischen Institut ETH, Stiftung Rubel, Zurich (1986)); vegetables including tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca sativa*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus* spp.), and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals include azalea (*Rhododendron* spp.), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa* spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum. Conifers that may be employed in practicing the present invention include, for example, pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*), Douglas-fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga canadensis*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as silver fir (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow-cedar (*Chamaecyparis nootkatensis*). Leguminous plants include beans and peas. Beans include guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, etc. Legumes include, but are not limited to, *Arachis*, e.g., peanuts, *Vicia*, e.g., crown vetch, hairy vetch, adzuki bean, mung bean, and chickpea, *Lupinus*, e.g., lupine, trifolium, *Phaseolus*, e.g., common bean and lima bean, *Pisum*, e.g., field bean, *Melilotus*, e.g., clover, *Medicago*, e.g., alfalfa, Lotus, e.g., trefoil, lens, e.g., lentil, and false indigo, *Acacia*, aneth, artichoke, arugula, blackberry, canola, cilantro, clementines, escarole, eucalyptus, fennel, grapefruit, honey dew, jicama, kiwifruit, lemon, lime, mushroom, nut, okra, orange, parsley, persimmon, plantain, pomegranate, poplar, radiata pine, radicchio, Southern pine, sweetgum, tangerine, triticale, vine, yams, apple, pear, quince, cherry, apricot, melon, hemp, buckwheat, grape, raspberry, chenopodium, blueberry, nectarine, peach, plum, strawberry, watermelon, eggplant, pepper, caluliflower, Brassica, e.g.,

broccoli, cabbage, brussels sprouts, onion, carrot, leek, beet, broad bean, celery, radish, pumpkin, endive, gourd, garlic, snapbean, spinach, squash, turnip, asparagus, and zucchini and ornamental plants include impatiens, Begonia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Agertum, Amaranthus, Antihirrhinum, Aquilegia, Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossos, and Zinnia. Other vegetables are in Table 1.

Preferred forage and turf grass for use in the methods of the invention include alfalfa, orchard grass, tall fescue, perennial ryegrass, creeping bent grass, and redtop.

Preferably, plants or cells to be transformed are crop plants and in particular cereals (for example, corn, alfalfa, sunflower, rice, *Brassica*, canola, soybean, barley, soybean, sugarbeet, cotton, safflower, peanut, sorghum, wheat, millet, tobacco, and the like), and even more preferably rice, corn and soybean.

In a preferred embodiment, the transformed host cells are monocot or dicot cells, including, but not limited to, wheat, corn (maize), rice, oat, barley, millet, rye, rape and alfalfa, as well as asparagus, tomato, egg plant, apple, pear, quince, cherry, apricot, pepper, melon, lettuce, cauliflower, *Brassica*, e.g., broccoli, cabbage, brussels sprout, sugar beet, sugar cane, sweetcorn, onion, carrot, leek, cucumber, tobacco, aubergine, beet, broad bean, carrot, celery, chicory, cotton, radish, pumpkin, hemp, buckwheat, orchardgrass, creeping bent top, redtop, ryegrass, tobacco, turfgrass, tall fescue, cow pea, endive, gourd, grape, raspberry, chenopodium, blueberry, pineapple, avocado, mango, banana, groundnut, nectarine, papaya, garlic, pea, peach, peanut, pepper, pineapple, plum, potato, safflower, snap bean, spinach, squashes, strawberry, sunflower, sorghum, sweet potato, turnip, watermelon, legumes such as *Arachis*, e.g., peanuts, *Vicia*, e.g., crown vetch, hairy vetch, adzuki bean, mung bean, and chickpea, *Lupinus*, e.g., lupine, trifolium, *Phaseolus*, e.g., common bean and lima bean, *Pisum*, e.g., field bean, *Melilotus*, e.g., clover, *Medicago*, e.g., alfalfa, Lotus, e.g., trefoil, lens, e.g., lentil, and false indigo, and the like; and ornamental crops including Impatiens, Begonia, Petunia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Ageratum, Amaranthus, Anthirrhinum, Aquilegia, Chrysanthemum, Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossis, Zinnia, and the like. More preferably, the transformed host cells are monocot cells such as maize, rice, wheat, barley, oats, and sorghum, which can be regenerated into a transgenic plant.

Any plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a vector of the present invention. The term "organogenesis," as used herein, means a process by which shoots and roots are developed sequentially from meristematic centers; the term "embryogenesis," as used herein, means a process by which shoots and roots develop together in a concerted fashion (not sequentially), whether from somatic cells or gametes. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons, hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (e.g., apical meristems, axillary buds, and root meristems), and induced meristem tissue (e.g., cotyledon meristem and hypocotyl meristem).

The choice of plant tissue source for transformation will depend on the nature of the host plant and the transformation protocol. Useful tissue sources include callus, suspension culture cells, protoplasts, leaf segments, stem segments, tassels, pollen, embryos, hypocotyls, tuber segments, meristematic regions, and the like. The tissue source is selected and transformed so that it retains the ability to regenerate whole, fertile plants following transformation, i.e., contains totipotent cells. Type I or Type II embryonic maize callus and immature embryos are preferred *Zea mays* tissue sources. Selection of tissue sources for transformation of monocots is described in PCT publication WO 95/06128.

For certain plant species, different antibiotic or herbicide selection markers may be preferred. Selection markers used routinely in transformation include the nptII gene which confers resistance to kanamycin and related antibiotics (Messing & Vierra, 1982); Bevan et al., 1983), the *bar* gene which confers resistance to the herbicide phosphinothricin (White et al., 1990, Spencer et al., 1990), the *hph* gene which confers resistance to the antibiotic hygromycin (Blochinger & Diggelmann), and the *dhfr* gene, which confers resistance to methotrexate (Bourouis et al., 1983).

Thus, the present invention also provides a transformed (transgenic) plant cell, *in planta* or *ex planta*, including, but not limited to, a transformed plant cell from plants such as corn (*Zea mays*), *Brassica* sp. (e.g., *B. napus*, *B. rapa*, *B. juncea*), particularly those *Brassica* species useful as sources of seed oil, alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), millet (e.g., pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*)), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat

(*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium barbadense, Gossypium hirsutum*), sweet potato (*Ipomoea batatus*), cassava (*Manihot esculenta*), coffee (*Cofea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea americana*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, barley, vegetables, ornamentals, and conifers; duckweed (*Lemna*, see WO 00/07210, which includes members of the family *Lemnaceae*. There are known four genera and 34 species of duckweed as follows: genus *Lemna (L. aequinoctialis, L. disperma, L. ecuadoriensis, L. gibba, L. japonica, L. minor, L. miniscula, L. obscura, L. perpusilla, L. tenera, L.* trisulca, L. turionifera, L. valdiviana); genus *Spirodela (S. intermedia, S. polyrrhiza, S.punctata);* genus *Woffia (Wa. angusta, Wa. arrhiza, Wa. australina, Wa. borealis, Wa. brasiliensis, Wa. columbiana, Wa. elongata, Wa. globosa, Wa. microscopica, Wa. neglecta)* and genus *Wofiella (Wl. caudata, Wl. denticulata, Wl. gladiata, Wl. hyalina, Wl. lingulata, Wl. repunda, Wl. rotunda, and Wl. neotropica*). Any other genera or species of *Lemnaceae*, if they exist, are also aspects of the present invention. *Lemna gibba, Lemna minor*, and *Lemna miniscula* are preferred, with *Lemna* minor and *Lemna miniscula* being most preferred. *Lemna* species can be classified using the taxonomic scheme described by Landolt, Biosystematic Investigation on the Family of Duckweeds: The family of Lemnaceae - A Monograph Study. Geobatanischen Institut ETH, Stiftung Rubel, Zurich (1986)); vegetables including tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca sativa*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus* spp.), and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals include azalea (*Rhododendron* spp.), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa* spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum. Conifers that may be employed in practicing the present invention include, for example, pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*), Douglas-fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga canadensis*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as

silver fir (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow-cedar (*Chamaecyparis nootkatensis*). Leguminous plants include beans and peas. Beans include guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, etc. Legumes include, but are not limited to, *Arachis*, e.g., peanuts, *Vicia*, e.g., crown vetch, hairy vetch, adzuki bean, mung bean, and chickpea, *Lupinus*, e.g., lupine, trifolium, *Phaseolus*, e.g., common bean and lima bean, *Pisum*, e.g., field bean, *Melilotus*, e.g., clover, *Medicago*, e.g., alfalfa, Lotus, e.g., trefoil, lens, e.g., lentil, and false indigo, *Acacia*, aneth, artichoke, arugula, blackberry, canola, cilantro, clementines, escarole, eucalyptus, fennel, grapefruit, honey dew, jicama, kiwifruit, lemon, lime, mushroom, nut, okra, orange, parsley, persimmon, plantain, pomegranate, poplar, radiata pine, radicchio, Southern pine, sweetgum, tangerine, triticale, vine, yams, apple, pear, quince, cherry, apricot, melon, hemp, buckwheat, grape, raspberry, chenopodium, blueberry, nectarine, peach, plum, strawberry, watermelon, eggplant, pepper, caluliflower, Brassica, e.g., broccoli, cabbage, brussels sprouts, onion, carrot, leek, beet, broad bean, celery, radish, pumpkin, endive, gourd, garlic, snapbean, spinach, squash, turnip, asparagus, and zucchini and ornamental plants include impatiens, Begonia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Agertum, Amaranthus, Antihirrhinum, Aquilegia, Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossos, and Zinnia, as well as from vegetables including those described in Table 1.

In a preferred embodiment, the transformed plants, include, but are not limited to, transformed wheat, corn (maize), rice, oat, barley, millet, rye, rape and alfalfa, as well as asparagus, tomato, egg plant, apple, pear, quince, cherry, apricot, pepper, melon, lettuce, cauliflower, *Brassica*, e.g., broccoli, cabbage, brussels sprout, sugar beet, sugar cane, sweetcorn, onion, carrot, leek, cucumber, tobacco, aubergine, beet, broad bean, carrot, celery, chicory, cotton, radish, pumpkin, hemp, buckwheat, orchardgrass, creeping bent top, redtop, ryegrass, tobacco, turfgrass, tall fescue, cow pea, endive, gourd, grape, raspberry, chenopodium, blueberry, pineapple, avocado, mango, banana, groundnut, nectarine, papaya, garlic, pea, peach, peanut, pepper, pineapple, plum, potato, safflower, snap bean, spinach, squashes, strawberry, sunflower, sorghum, sweet potato, turnip, watermelon, legumes such as *Arachis*, e.g., peanuts, *Vicia,* e.g., crown vetch, hairy vetch, adzuki bean, mung bean, and chickpea, *Lupinus*, e.g., lupine, trifolium, *Phaseolus*, e.g., common bean and lima bean, *Pisum,* e.g., field bean, *Melilotus*, e.g., clover, *Medicago*, e.g., alfalfa, Lotus, e.g., trefoil, lens, e.g.,

lentil, and false indigo, and the like; and ornamental crops including Impatiens, Begonia, Petunia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Ageratum, Amaranthus, Anthirrhinum, Aquilegia, Chrysanthemum, Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossis, Zinnia, and the like. Preferably, the transformed plants are transformed monocot such as maize, rice, wheat, barley, oats, and sorghum.

IV. Identification of Transgenic Plants

To confirm the presence of the preselected nucleic acid segment(s) or "transgene(s)" in the regenerating plants, a variety of assays may be performed. Such assays include, for example, "molecular biological" assays well known to those of skill in the art, such as Southern and Northern blotting, *in situ* hybridization and nucleic acid-based amplification methods such as PCR or RT-PCR; "biochemical" assays, such as detecting the presence of a protein product, e.g., by immunological means (ELISAs and Western blots) or by enzymatic function; plant part assays, such as leaf or root assays; and also, by analyzing the phenotype of the whole regenerated plant, e.g., for disease or pest resistance.

DNA may be isolated from cell lines or any plant parts to determine the presence of the preselected nucleic acid segment through the use of techniques well known to those skilled in the art. Note that intact sequences will not always be present, presumably due to rearrangement or deletion of sequences in the cell.

The presence of nucleic acid elements introduced through the methods of this invention may be determined by polymerase chain reaction (PCR). Using this technique discreet fragments of nucleic acid are amplified and detected by gel electrophoresis. This type of analysis permits one to determine whether a preselected nucleic acid segment is present in a stable transformant, but does not prove integration of the introduced preselected nucleic acid segment into the host cell genome. In addition, it is not possible using PCR techniques to determine whether transformants have exogenous genes introduced into different sites in the genome, i.e., whether transformants are of independent origin. It is contemplated that using PCR techniques it would be possible to clone fragments of the host genomic DNA adjacent to an introduced preselected DNA segment.

Positive proof of DNA integration into the host genome and the independent identities of transformants may be determined using the technique of Southern hybridization. Using this technique specific DNA sequences that were introduced into the host genome and flanking

host DNA sequences can be identified. Hence the Southern hybridization pattern of a given transformant serves as an identifying characteristic of that transformant. In addition it is possible through Southern hybridization to demonstrate the presence of introduced preselected DNA segments in high molecular weight DNA, i.e., confirm that the introduced preselected DNA segment has been integrated into the host cell genome. The technique of Southern hybridization provides information that is obtained using PCR, e.g., the presence of a preselected DNA segment, but also demonstrates integration into the genome and characterizes each individual transformant.

It is contemplated that using the techniques of dot or slot blot hybridization which are modifications of Southern hybridization techniques one could obtain the same information that is derived from PCR, e.g., the presence of a preselected DNA segment.

Both PCR and Southern hybridization techniques can be used to demonstrate transmission of a preselected DNA segment to progeny. In most instances the characteristic Southern hybridization pattern for a given transformant will segregate in progeny as one or more Mendelian genes (Spencer et al., 1992); Laursen et al., 1994) indicating stable inheritance of the gene. The nonchimeric nature of the callus and the parental transformants (R_0) was suggested by germline transmission and the identical Southern blot hybridization patterns and intensities of the transforming DNA in callus, R_0 plants and R_1 progeny that segregated for the transformed gene.

Whereas DNA analysis techniques may be conducted using DNA isolated from any part of a plant, RNA may only be expressed in particular cells or tissue types and hence it will be necessary to prepare RNA for analysis from these tissues. PCR techniques may also be used for detection and quantitation of RNA produced from introduced preselected DNA segments. In this application of PCR it is first necessary to reverse transcribe RNA into DNA, using enzymes such as reverse transcriptase, and then through the use of conventional PCR techniques amplify the DNA. In most instances PCR techniques, while useful, will not demonstrate integrity of the RNA product. Further information about the nature of the RNA product may be obtained by Northern blotting. This technique will demonstrate the presence of an RNA species and give information about the integrity of that RNA. The presence or absence of an RNA species can also be determined using dot or slot blot Northern hybridizations. These techniques are modifications of Northern blotting and will only demonstrate the presence or absence of an RNA species.

While Southern blotting and PCR may be used to detect the preselected DNA segment in question, they do not provide information as to whether the preselected DNA segment is being expressed. Expression may be evaluated by specifically identifying the protein products of the introduced preselected DNA segments or evaluating the phenotypic changes brought about by their expression.

Assays for the production and identification of specific proteins may make use of physical-chemical, structural, functional, or other properties of the proteins. Unique physical-chemical or structural properties allow the proteins to be separated and identified by electrophoretic procedures, such as native or denaturing gel electrophoresis or isoelectric focussing, or by chromatographic techniques such as ion exchange or gel exclusion chromatography. The unique structures of individual proteins offer opportunities for use of specific antibodies to detect their presence in formats such as an ELISA assay. Combinations of approaches may be employed with even greater specificity such as Western blotting in which antibodies are used to locate individual gene products that have been separated by electrophoretic techniques. Additional techniques may be employed to absolutely confirm the identity of the product of interest such as evaluation by amino acid sequencing following purification. Although these are among the most commonly employed, other procedures may be additionally used.

Assay procedures may also be used to identify the expression of proteins by their functionality, especially the ability of enzymes to catalyze specific chemical reactions involving specific substrates and products. These reactions may be followed by providing and quantifying the loss of substrates or the generation of products of the reactions by physical or chemical procedures. Examples are as varied as the enzyme to be analyzed.

Very frequently the expression of a gene product is determined by evaluating the phenotypic results of its expression. These assays also may take many forms including but not limited to analyzing changes in the chemical composition, morphology, or physiological properties of the plant. Morphological changes may include greater stature or thicker stalks. Most often changes in response of plants or plant parts to imposed treatments are evaluated under carefully controlled conditions termed bioassays.

V. Utility

Once an expression cassette of the invention has been transformed into a particular plant species, it may be propagated in that species or moved into other varieties of the same species,

particularly including commercial varieties, using traditional breeding techniques. Particularly preferred plants of the invention include the agronomically important crops listed above. The genetic properties engineered into the transgenic seeds and plants described above are passed on by sexual reproduction and can thus be maintained and propagated in progeny plants. The present invention also relates to a transgenic plant cell, tissue, organ, seed or plant part obtained from the transgenic plant. Also included within the invention are transgenic descendants of the plant as well as transgenic plant cells, tissues, organs, seeds and plant parts obtained from the descendants.

Preferably, the expression cassette in the transgenic plant is sexually transmitted. In one preferred embodiment, the coding sequence is sexually transmitted through a complete normal sexual cycle of the R0 plant to the R1 generation. Additionally preferred, the expression cassette is expressed in the cells, tissues, seeds or plant of a transgenic plant in an amount that is different than the amount in the cells, tissues, seeds or plant of a plant which only differs in that the expression cassette is absent.

The transgenic plants produced herein are thus expected to be useful for a variety of commercial and research purposes. Transgenic plants can be created for use in traditional agriculture to possess traits beneficial to the grower (e.g., agronomic traits such as resistance to water deficit, pest resistance, herbicide resistance or increased yield), beneficial to the consumer of the grain harvested from the plant (e.g., improved nutritive content in human food or animal feed), or beneficial to the food processor (e.g., improved processing traits). In such uses, the plants are generally grown for the use of their grain in human or animal foods. However, other parts of the plants, including stalks, husks, vegetative parts, and the like, may also have utility, including use as part of animal silage or for ornamental purposes. Often, chemical constituents (e.g., oils or starches) of maize and other crops are extracted for foods or industrial use and transgenic plants may be created which have enhanced or modified levels of such components.

Transgenic plants may also find use in the commercial manufacture of proteins or other molecules, where the molecule of interest is extracted or purified from plant parts, seeds, and the like. Cells or tissue from the plants may also be cultured, grown *in vitro*, or fermented to manufacture such molecules.

The transgenic plants may also be used in commercial breeding programs, or may be crossed or bred to plants of related crop species. Improvements encoded by the expression

cassette may be transferred, e.g., from maize cells to cells of other species, e.g., by protoplast fusion.

The transgenic plants may have many uses in research or breeding, including creation of new mutant plants through insertional mutagenesis, in order to identify beneficial mutants that might later be created by traditional mutation and selection. An example would be the introduction of a recombinant DNA sequence encoding a transposable element that may be used for generating genetic variation. The methods of the invention may also be used to create plants having unique "signature sequences" or other marker sequences which can be used to identify proprietary lines or varieties.

Thus, the transgenic plants and seeds according to the invention can be used in plant breeding which aims at the development of plants with improved properties conferred by the expression cassette, such as tolerance of viruses or other pests, or other stresses. The various breeding steps are characterized by well-defined human intervention such as selecting the lines to be crossed, directing pollination of the parental lines, or selecting appropriate descendant plants. Depending on the desired properties different breeding measures are taken. The relevant techniques are well known in the art and include but are not limited to hybridization, inbreeding, backcross breeding, multiline breeding, variety blend, interspecific hybridization, aneuploid techniques, etc. Hybridization techniques also include the sterilization of plants to yield male or female sterile plants by mechanical, chemical or biochemical means. Cross pollination of a male sterile plant with pollen of a different line assures that the genome of the male sterile but female fertile plant will uniformly obtain properties of both parental lines. Thus, the transgenic seeds and plants according to the invention can be used for the breeding of improved plant lines which for example increase the effectiveness of conventional methods such as herbicide or pesticide treatment or allow to dispense with said methods due to their modified genetic properties. Alternatively new crops with improved stress tolerance can be obtained which, due to their optimized genetic "equipment", yield harvested product of better quality than products which were not able to tolerate comparable adverse developmental conditions.

The invention will be further described by the following examples which is not intended to limit the scope of the invention.

Example 1

To identify novel genes that are associated with the high protein phenotype in selected lines of *Zea mays*, a differential analysis of four high protein lines and two control lines as well as a segregating population derived from a high protein line and a normal line was conducted. High protein corn lines (Wil500, Wil578, WIO465), control lines (WICY530 and LH59) and a segregating population derived from cross LH59XWIL578 (total of 53 lines) were obtained from Wilson Genetics. High protein maize refers to germplasm having elevated levels of protein in the seed, typically above 14.5 % in the whole kernel, above 17% in the embryo and above 13.5% in the endosperm (see Figure 1).

The following proteomic approaches were used: 1) extraction of proteins from tissue including total kernels, mature/developing embyros, root and leaf from 2 week old seedlings, optionally exposed to fertilizer; 2) two-dimensional (2-D) separation of proteins by size and charge using gels using isoelectric focusing (IEF) and SDS-PAGE at three pH ranges (e.g., pH 3-10, pH 4-7 and pH 7-10); 3) image analysis of silver stained gels to identify differentially expressed proteins (by visual inspection and PDQUEST software); 4) gel excision and trypsin digestion of selected protein spots; 5) analysis of resulting tryptic peptides using MALDI-TOF mass spectrometry; and 6) database searching using protein sequence information for protein identification using SEQUEST.

Materials and Methods

Sample preparation and gel electrophoresis. Embryos from mature corn kernel from both high protein lines and normal corn lines were cut out of the seeds and directly homogenized in a solution containing 7 uM urea, 2 uM thiourea, 0.5% Triton X-100 and 60 mM DTT. The first dimension for isoelectric focusing was carried out on a BioRad IPG system essentially as described by the manufacturer
using three pH gradient strips, pH 3-10, pH 4-7 and pH 5-8 for 45kvhr. Subsequent to loading the IEF strips on the second dimension, the IEF strips were re-equilibrated with a solution (2% SDS, 50 mM Tris, pH 6.9, 10% glycerol and 7 mM urea), and directly applied to a BioRad 8-16% gradient SDS-PAGE gel for electrophoresis. The resultant gels were stained with silver using a BioRad silver staining kit according to the manufacturer's recommendations. 2D PAGE profiles were laser scanned and comparative analyses were performed using PDQuest

software package (BioRad) . Only spots that were present/completely absent between normal and high protein lines were selected for further analysis. Protein spots were cut out of the gel either manually or using the BioRad spot cutter.

Trypsin digestion. Gel pieces were transferred to an eppendorf tube or a polypropylene 96 well plate. 100 ul acetonitrile was added to dehydrate the gel. After removing the acetonitrile by speed vacuum, the gels were contacted with 50 mM NH_4HCO_3 and trypsin at 10 ng/ul and digested overnight at 37 degrees C. Peptides were extracted by 3 washes with 5% formic acid in 50% acetonitrile. The combined supernatants were dried down in a Speedvac and the peptides were redissolved in 6 ul of 0.1% formic acid for MS analysis.

MS/MS analysis and data analysis. All analysis were performed on a Finnigan LCQ ion trap mass spectrometer that was run and operated as described in Link et al. (1997). The peptide sequence raw data was searched against a cereal database by SEQUEST software. To determine the function of the genes identified as being differentially expressed, a number of criteria were considered: the statistical score from SEQUEST, xcorr and deltCN, the peptide length and terminal sequence, the quality of the spectrum from the peptides, the number of peptides from the same protein spots that were identified in the same search, and the molecular weight and pI of the protein.

Results

Approximately 100 2-D gels under three different conditions were analyzed with samples from 2 normal lines, 3 high protein lines and 10 selected lines from the segregating population. Using data from mature embryos from Wil500, WIO465, LH59 and WICY530, approximately 120 differentially expressed proteins in were ideneified and isolated. Figure 2 is an example of two gels, one with proteins from control maize embryos (pH 5-8, spots 13-18, panel A) and another with proteins from a high protein line (pH 5-8, spots 1-12, panel B). Figures 2C and 2D are further examples, in which the arrow points to a readily identifiable difference area that contains the various forms of globulins proteins in embryo as described in the invention.

Thirty-eight of the differentially expressed maize proteins or their orthologs are listed in the Sequence Listing. For example, the following proteins were found to be differentially expressed: globulin-1 s allele precursor, globulin 2 precursor, glucose and ribitol dehydrogenase, glutathione S-transferase, rab28 protein (maize), heat shock protein 17.2, oleosin 16 kD protein, and putative receptor protein kinase zmpk1 precursor. In general, there was more globulin in all the high protein lines tested and there appears to be a very different

mature product of the glb1 and glb2 genes in high protein lines. These differences may occur due to regulatory processes, allelic variation, at the mRNA and/or protein level or post-translationally. Figures 3A to 3H show a subset of the 38 proteins which were identified using different criteria (Xcorr and dCN).

This genetic information is useful for marker development for breeding purposes in maize and for seed protein content manipulation in cereals in general.

Example 2

Gene expression levels of three of the genes of the invention, with sequences in the Sequence Listing, were examined during seed development in rice using microarray technology. Relative gene expression levels were determined and are presented in Table 2. All three genes were up regulated in the time course of rice seed development. The gene expression levels were determined by hybridizing the rice mRNA isolated at various developmental stages to an Affymetrix gene chip containing rice gene sequences. The rice genechip covered about 20,000 rice genes. A similar pattern of gene experession during corn seed maturation si expected.

Table 2

Developmetal stages	17 kd Heat shock	Dehydrin	GRD
seed development f_Seed day 0 anthesis	50.08	35.49	117.72
seed development g_Seed day 2 post anthesis	131.87	36.5	132.49
seed development h_Seed day 4 post anthesis	193.08	45.24	149.95
seed development I_Seed day 7 post anthesis	383.84	151.97	325.3
seed development j_Seed day 9 post anthesis	465.35	194.87	426.05
seed development k_11-day_post anthesis	737.56	356.98	784.33
seed development l_14-day_post anthesis	632.7	457.08	840.55
seed development m_17-day_post anthesis	1,150.06	438.75	1,486.17
seed development n_19-day_post anthesis	1,268.10	433.28	1,702.69

Example 4

Co-segregation of 17kD heat shock protein gene expression with high protein phenotype was observed. Tagman analysis (real time PCR) of 17 kD heat shock protein gene expression was compared with unbiqutin gene expression level as a reference. 30DAP embryos were used in these experiments. PCR primers were designed according the sequences shown in the Sequence Listing. This result demonstrates the cosegregation of HS17 gene expression with the high protein phenotype in the hybrids tested. These are the same corn lines as showed in Figure 1.

Table 3

Corn lines	HS 17 gene	UBQ gene
73/93 High Protein	1.59	1
73/98 High Protein	0.69	1
73/88 High Protein	5.05	1
73/92 Low Protein	0.07	1
73/76 Low Protein	0.01	1
73/84 Low Protein	0.07	1
LH59 Normal Protein	0.29	1
WIL500 high protein	5.33	1

Taqman analysis for key candidate gene expression was performed as follows. For one step RT-PCR amplification, total RNA was used in a 50 ml reaction using the master mixture of a Taq-Man One-Step RT-PCR Mix Reagnets (cat # 4309169, lot# 0006014) (PE Biosystems, Foster City, CA), following the manufacturer's protocol. The one step RT-PCR was conducted with an ABI Prism* 7900 HT Sequence Detection System (AB Applied Biosystems, Foster City, CA). The reactions were incubated for 30 min at 48° C for reverse transcription, and for 40 cycles of 15 s at 95° C, 60 s at 60° C for amplification. The ramp rate was set at 100%

between two different temperature set points. 50 ml Reaction was composed of 6.25 ml of 2 mM forward primer, AtTRX3-F (gtgtggaaatgacacagattgtga), 6.25 ml of 2 mM reverse primer, AtTRX3-R (agacgggtgcaatgaaacg), 6.25 ml of 2 mM TaqMan probe (6FAM-agacttcactgcaacatggtgcccac-TAMRA), AtTRX3_TaqMan, 1.25 ml of 40x MultiScribe & Rnase inhibitor Mix, 5 ml of template RNA (50 ng total RNA), and 25 ml of Master Mix w/o UNG (Taq-Man One-Step RT-PCR Mix Reagnets: cat # 4309169, lot# 0006014) (PE Biosystems, Foster City, CA). Data collection was processed between two temperature set points of 95° C and 60° C during amplification. The fold change in TRX3 transcript was determined following the ABI Prism 7900HT Sequence Detection System User Guide (Applied Biosystems): Fold change = 2-DCt, where DCt = -(Ct TRX - CtTRX-STD1ng) threshold 0.36507.

Example 5

Modulation of high protein trait by genes of the invention is readily determined using plant transformation sytems as described herein and as known in the art. In one embodiment, the Gateway cloning system was used to introduce genes of the invention into agrotransformation vectors for cereals, with seed specific promoters. See Figures 4A and 4B. The embryo specific promoter is a globulin promoter, and the ADPGPP gene promoter is used as the endosperm specific promoter. Use of these promoter constructs allows ease of cloning various genes under the control of these promoter to overexpress and/or downregulate the expression of these genes.

Gateway System cloning of pOPT003 & pOPT004 was as follows. Two oligos (NJ001 for & NJ002rev) were designed to amplify the Gateway Cassette A. These oligos contain restriction enzymes (Bcl I and Spe I) to clone into Xba I and BamH I sites of the pNOV4000 and pNOV4002 vector (note that Xba I is compatible with Spe I site and Bcl I is compatible with BamH I site). The sGFP-M5 gene of the pNOV4000 and pNOV4002 plasmid is replaced with the Gateway cassette A in which we generated pOPT001 and pOPT002 vector. pOPT001, pOPT002 and pNOV2117 (agro) vector were digested and ligated with Kpn I and Hind III sites. The final products were transformed into DB3.1 E.Coli cells, and the pOPT003 and pOPT004 vectors were generated, as shown in the figures.

Protein determination was done as follows. For green house generated materials, seed protein were determined by elemental analysis, nitrogen to caculate the total protein yield, using conversion factor 6.25. A N/protein analyzer, FLASH EA 1112 Series , from CE

Instruments were used in our experiment. Protein content for field generated seed materials were determined by NIR (Near Infrared) analysis.

Example 6

Vector construction for overexpression and gene "knockout" experiments.

Overexpression

Vectors used for expression of full-length genes of the embodiments of the invention of interest in plants (overexpression) are designed to overexpress the protein of interest and are of two general types, biolistic and binary, depending on the plant transformation method to be used.

For biolistic transformation (biolistic vectors), the requirements are as follows:

1. a backbone with a bacterial selectable marker (typically, an antibiotic resistance gene) and origin of replication functional in *Escherichia coli* (*E. coli*; eg. ColE1), and
2. a plant-specific portion consisting of:
    a. a gene expression cassette consisting of a promoter (eg. ZmUBIint MOD), the gene of interest (typically, a full-length cDNA) and a transcriptional terminator (eg. *Agrobacterium tumefaciens nos* terminator);
    b. a plant selectable marker cassette, consisting of a promoter (eg. rice Act1D-BV MOD), selectable marker gene (eg. phosphomannose isomerase, PMI) and transcriptional terminator (eg. CaMV terminator).

Vectors designed for transformation by *Agrobacterium tumefaciens* (*A. tumefaciens*; binary vectors) consist of:

1. a backbone with a bacterial selectable marker functional in both *E. coli* and *A. tumefaciens* (eg. spectinomycin resistance mediated by the *aadA* gene) and two origins of replication, functional in each of aforementioned bacterial hosts, plus the *A. tumefaciens virG* gene;
2. a plant-specific portion as described for biolistic vectors above, except in this instance this portion is flanked by *A. tumefaciens* right and left border sequences which mediate transfer of the DNA flanked by these two sequences to the plant.

Knock out vectors

Vectors designed for reducing or abolishing expression of a single gene or of a family or related genes (knockout vectors) are also of two general types corresponding to the

methodology used to downregulate gene expression: antisense or double-stranded RNA interference (dsRNAi).

Anti-sense

For antisense vectors, a full-length or partial gene fragment (typically, a portion of the cDNA) can be used in the same vectors described for full-length expression, as part of the gene expression cassette. For antisense-mediated down-regulation of gene expression, the coding region of the gene or gene fragment will be in the opposite orientation relative to the promoter; thus, mRNA will be made from the non-coding (antisense) strand *in planta*.

dsRNAi

For dsRNAi vectors, a partial gene fragment (typically, 300 to 500 basepairs long) is used in the gene expression cassette, and is expressed in both the sense and antisense orientations, separated by a spacer region (typically, a plant intron, eg. the OsSH1 intron 1, or a selectable marker, eg. conferring kanamycin resistance). Vectors of this type are designed to form a double-stranded mRNA stem, resulting from the basepairing of the two complementary gene fragments *in planta*.

Biolistic or binary vectors designed for overexpression or knockout can vary in a number of different ways, including eg. the selectable markers used in plant and bacteria, the transcriptional terminators used in the gene expression and plant selectable marker cassettes, and the methodologies used for cloning in gene or gene fragments of interest (typically, conventional restriction enzyme-mediated or Gateway™ recombinase-based cloning). An important variant is the nature of the gene expression cassette promoter driving expression of the gene or gene fragment of interest in most tissues of the plants (constitutive, eg. ZmUBIint MOD), in specific plant tissues (eg. maize ADP-gpp for endosperm-specific expression), or in an inducible fashion (eg. GAL4bsBz1 for estradiol-inducible expression in lines constitutively expressing the cognate transcriptional activator for this promoter).

Insertion of a gene of the embodiments of the invention into Expression Vector

A validated rice cDNA clone such as the *OsPT11* cDNA prepared in Example 14 above, in pCR2.1-TOPO is subcloned using conventional restriction enzyme-based cloning into a vector, downstream of the maize ubiquitin promoter and intron, and upstream of the *Agrobacterium tumefaciens* nos 3' end transcriptional terminator. The resultant gene expression cassette (promoter, gene of the embodiments of the invention and terminator) is

further subcloned, using conventional restriction enzyme-based cloning, into the pNOV2117 binary vector, generating pNOVCAND.

The pNOVCAND binary vector is designed for transformation and over-expression of the gene of the embodiments of the invention in monocots. It consists of a binary backbone containing the sequences necessary for selection and growth in *Escherichia coli* DH-5α (Invitrogen) and *Agrobacterium tumefaciens* LBA4404, including the bacterial spectinomycin antibiotic resistance *aadA* gene from *E. coli* transposon Tn7, origins of replication for *E. coli* (ColE1) and *A. tumefaciens* (VS1), and the *A. tumefaciens virG* gene. In addition to the binary backbone, pNOV2117 contains the T-DNA portion flanked by the right and left border sequences, and including the Positech™ (Syngenta) plant selectable marker and the gene of the embodiments of the invention gene expression cassette. The Positech™ plant selectable marker confers resistance to mannose and in this instance consists of the maize ubiquitin promoter driving expression of the PMI (phosphomannose isomerase) gene, followed by the cauliflower mosaic virus transcriptional terminator.

This is exemplified in Rice Transformation as follows. pNOVCAND is transformed into a rice cultivar (Kaybonnet) using Agrobacterium-mediated transformation, and mannose-resistant calli are selected and regenerated.

Agrobacterium is grown on YPC solid plates for 2-3 days prior to experiment initiation. Agrobacterial colonies are suspended in liquid MS media to an OD of 0.2 at λ600nm. Acetosyringone is added to the agrobacterial suspension to a concentration of 200μM and agro is induced for 30min.

Three-week-old calli which are induced from the scutellum of mature seeds in the N6 medium (Chu, C.C. et al., Sci, Sin., 18, 659-668(1975)) are incubated in the agrobacterium solution in a 100 x 25 petri plate for 30 minutes with occasional shaking. The solution is then removed with a pipet and the callus transfered to a MSAs medium which is overlayed with sterile filter paper.

Co-Cultivation is continued for 2 days in the dark at 22°C.

Calli are then placed on MS-Timetin plates for 1 week. After that they are tranfered to PAA + mannose selection media for 3 weeks.

Growing calli (putative events) are picked and transfered to PAA+ mannose media and cultivated for 2 weeks in light.

Colonies are tranfered to MS20SorbKinTim regeneration media in plates for 2 weeks in light. Small plantlets are transferred to MS20SorbKinTim regeneration media in GA7 containers. When they reach the lid, they are transfered to soil in the greenhouse.

Expression of the gene of the embodiments of the invention in transgenic T_0 plants is analyzed. Additional rice cultivars, such as but not limited to, Nipponbare, Taipei 309 and Fuzisaka 2 are also transformed and assayed for expression of the gene product of the embodiments of the invention and enhanced protein expression.

References

Altschul et al., J. Mol. Biol. 215:403 (1990).
Altschul et al., Nucleic Acids Res. 25:3389 (1997).
An et al., EMBO J. 4:277 (1985).
Aoyama T. et al., N-H Plant Journal, 11:605 (1997).
Ballas et al., Nucleic Acids Res. 17:7891 (1989).
Bansal et al., Proc. Natl. Acad. Sci. USA, 89:3654 (1992).
Batzer et al., Nucleic Acid Res., 19:5081 (1991).
Beals et al., Plant Cell, 9:1527 (1997).
Belanger et al., Genetics, 129:863 (1991).
Bevan et al., Nature, 304:184 (1983).
Blochinger & Diggelmann, Mol Cell Biol, 4:2929.
Bourouis et al., EMBO J., 2:1099 (1983).
Brandle et al., Biotech. 13:994 (1995).
Byrne et al. Plant Cell Tissue and Organ Culture, 8:3 (1987).
Campbell and Gowri, Plant Physiol., 92:1 (1990).
Chandler et al., Plant Cell, 1:1175 (1989).
Christou et al., Biotechnology, 9:957 (1991).
Christou et al., Plant Physiol., 87:671 (1988).
Cordero et al., Plant J., 6:141 (1994).
Corpet et al., Nucleic Acids Res. 16:10881 (1988).
Crameri et al., Nature Biotech., 15:436 (1997).
Crameri et al., Nature, 391:288 (1998).

Crossway et al., BioTechniques, 4:320 (1986).

Czako et al., Mol. Gen. Genet. 23 5 (1), 33-40 (1992).

Datta et al., Bio/Technology, 8, 736 (1990).

Dayhoff et al., Atlas of Protein Sequence and Structure, Natl. Biomed. Res. Found., Washington, C.D. (1978).

De Blaere et al., Meth. Enzymol. 143:277 (1987).

de Framond, FEBS, 290:103 (1991).

Della-Cioppa et al., Plant Physiology, 84:965 (1987).

Dennis et al., Nucleic Acids Res., 12:3983 (1984).

Dzelkalns et al., Plant Cell, 5:855 (1993).

Elroy-Stein et al., PNAS USA, 86:6126 (1989).

English, et al., Plant Cell 8:179 (1996).

Eulgem et al., Trends in Plant Sci. 5:199 (2000).

Flavell, Proc. Natl. Acad. Sci USA 91:3490 (1994).

Franken et al., EMBO J., 10:2605 (1991).

Fromm et al., Bio/Technology 8:833 (1990).

Gallie et al., Molecular Biology of RNA, 237 (1989).

Gallie et al., Nucl. Acids Res., 15:8693 (1987).

Gan et al., Science (1995), 270 (5244), 1986-8).

Gatz Current Opinion in Biotechnology, 7:168 (1996).

Gatz, C., Annu. Rev. Plant Physiol. Plant Mol. Biol., 48:89 (1997).

Gelfand, eds., PCR Strategies (Academic Press, New York (1995)).

Gordon-Kamm et al., Plant Cell, 2, 603 (1990).

Graham et al., Biochem. Biophys. Res. Comm., 101:1164 (1981).

Graham et al., J. Biol. Chem., 260:6555 (1985).

Graham et al., J. Biol. Chem., 260:6561 (1985).

Guerineau et al., Mol. Gen. Genet. 262:141 (1991).

Hansel et al., Cereal Chem. 50:383 (1973).

Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA, 89:10915 (1989).

Hiei et al., Plant J. 6:271 (1994).

Higgins et al., Gene 73:237 (1988).

Higgins et al., CABIOS 5:151 (1989).

Hinchee et al., Biotechnology, 6:915 (1988).

Hoekema, In: The Binary Plant Vector System. Offset-drukkerij Kanters B.V.

Huang et al., CABIOS 8:155 (1992).

Hudspeth & Grula, Plant Molec. Biol., 12:579 (1989).

Huffman et al., J. Cell. Biochem., 17B: Abstract.

Ingelbrecht et al., Plant Cell, 1:671 (1989).

Innis et al., eds., PCR Protocols: A Guide to Methods and Applications (Academic Press, New York (1995).

Innis and Gelfand, eds., PCR Methods Manual (Academic Press, New York) (1999).

Jobling et al., Nature, 325:622 (1987).

John et al., Proc. Natl. Acad. Sci. USA 89(13):5769 (1992).

Jones, Biochem. Genet. 16:27 (1978).

Jones et al., Adv Bot Res 24:89 (1997).

Joshi et al., Nucleic Acid Res. 15:9627 (1987).

Joshi, NAR, 15:6643 (1987).

Karlin and Altschul, Proc. Natl. Acad Sci. USA 872264 (1990).

Karlin and Altschul, Proc. Natl. Acad. Sci. USA 90:5873 (1993).

Keller et al., Genes Dev., 3:1639 (1989).

Klein et al., Bio/Technology, 6:559 (1988).

Klein et al., Nature (London) 327:70 (1987).

Klein et al., Plant Physiol., 91:440 (1988).

Klein et al., Proc. Natl. Acad. Sci. USA, 85:4305 (1988).

Kohler et al., Plant Mol. Biol., 29:1293 (1995).

Knauf, et al., Genetic Analysis of Host Range Expression by *Agrobacterium* In: Molecular Genetics of the Bacteria-Plant Interaction, Puhler, A. ed., Springer-Verlag, New York, 245 (1983).

Koziel et al., Biotechnology, 11:194 (1993).

Kridl et al., Seed Science Research, 1:209 (1991).

Kriz et al., Mol. Gen. Genet., 207:90 (1987).

Kunkel, Proc. Natl. Acad. Sci. USA, 82:488 (1985).

Kunkel et al., Methods in Enzymol., 154:367 (1987).

Langridge et al., Cell, 34:1015 (1983).

Lindstrom et al., Der. Genet., 11:160 (1990).

Link et al., Electrophoresis, 18:1314 (1997).

Lommel et al., Virology, 81:382 (1991).

Lorz et.al. Mol. Gen. Genet. 199:178 (1985).

Macejak et al., Nature, 353:90 (1991).

Mansson et al., Gen. Genet., 200:356 (1985).

Martinez et al., J. Mol. Biol., 208:551 (1989).

Matzke and Matzke, Plant Physiol. 107:679 (1995).

McBride et al., Proc. Natl. Acad. Sci. USA, 91:7301 (1994).

McCabe et al., Bio/Technology, 6:923 (1988).

McDowell et al., Plant Cell 10:1861 (1998).

Meinkoth and Wahl, Anal. Biochem., 138:267 (1984).

Mertz et al., Science 145:279 (1964).

Messing & Vierra, Gene, 19:259 (1982).

Misra et al., Science 176:1425 (1972).

Mogen et al., Plant Cell 2:1261 (1990).

Moore et al., J. Mol. Biol., 272:336 (1997).

Munroe et al., Gene 91:151 (1990).

Murray et al., Nucleic Acids Res., 17:477 (1989).

Myers and Miller, CABIOS 4:11 (1988).

Needleman and Wunsch, J. Mol. Biol. 48:443 (1970).

Odell et al., Nature, 313:810 (1985).

Odell et al., Nature, 313:810 (1985).

Ohtsuka et al., J. Biol. Chem., 260:2605 (1985).

Okamuro et al., Biochemistry of Plants, 15:1 (1989).

Paszkowski et al., EMBO J., 3:2717 (1984).

Pacciotti et al. Bio/Technology 3:241 (1985).

Park et al., J. Plant Biol. 38(4):365 (1985).

Paulis et al., Cereal Chem. 55:705 (1978).

Pearson et al., Meth. Mol. Biol. 24:307 (1994).

Pearson and Lipman, Proc. Natl. Acad. Sci. 85:2444 (1988).

Penninck et al., Plant Cell 8:2309 (1996).

Perlak et al., Proc. Natl. Acad. Sci. USA, 88:3324 (1991).

Potrykus Mol. Gen. Genet. 199:183 (1985).

Proudfoot, Cell 64:671 (1991).

Quigley et al., J. Mol. Evol., 29:412 (1989).

Ralston et al., Genetics, 119:185 (1988).

Reina et al., Nucleic Acids Res., 18:6425 (1990).

Reina et al., Nucleic Acids Res., 18:7449 (1990).

Riggs et al., Proc. Natl. Acad. Sci. USA, 83:5602 (1986).

Rochester et al., (1986).

Rodermel et al., Cell 55:673 (1988).

Rossolini et al., Mol. Cell. Probes, 8:91 (1994).

Rubenstein, In: Maize for Biological Research (Sheridan ed., U of N. Dakota Press, pp 189-196 (1982)).

Ruiz et al., Plant Cell 10:937 (1998).

Sambrook et al., Molecular Cloning: A Laboratory Manual (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New York) (1989).

Sanfacon et al., Genes Dev. 5:141 (1991).

Sanford et al., Particulate Science and Technology, 5:27 (1987).

Schernthaner et al., EMBO J., 7:1249 (1988).

Schwob et al., Plant J., 4:423 (1993).

Shimamoto et al., Nature, 338:274 (1989).

Simpson, Plant Mol. Biol., 19:699 (1985).

Skuzeski et al., Plant Molec. Biol., 15:65 (1990).

Slater et al., Plant Mol. Biol., 5:137 (1985).

Smith et al., Adv. Appl. Math. 2:482 (1981).

Smith et al., Planta 168:94 (1986).

Spencer et al., Theor Appl Genet, 79:625 (1990).

Staub et al., EMBO J., 12:601 (1993).

Staub et al., Plant Cell, 4:39 (1992).

Stemmer, Nature, 370:389 (1994).

Stemmer, Proc. Natl. Acad. Sci. USA, 91:10747 (1994).

Sukhapinda et al. Plant Mol. Biol. 8:209 (1987).

Sullivan et al., Mol. Gen. Genet., 215:431 (1989).

Svab et al., Proc. Natl. Acad. Sci. USA, 87:8526 (1990).

Svab et al., Proc. Natl. Acad. Sci. USA, 90:913 (1993).

Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes, part I chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays" Elsevier, New York (1993).

Tsai et al., Biochem. Genet. 16:883 (1978).

Turner et al., Molecular Biotechnology, 3:225 (1995).

van der Krol et al., 1988a; 1988b, 1988c.

VanTunen et al., EMBO J., 7:1257 (1988).

Vasil et al., Biotechnology, 11:1553 (1993).

Vodkin, Prog. Clin. Biol. Res., 138:87 (1983).

Vogel et al., EMBO J., 11:157 (1992).

Walker and Gaastra, eds., Techniques in Molecular Biology, MacMillan Publishing Company, New York (1983).

Wandelt et al., Nucleic Acids Res., 17:2354 (1989).

Waterman, M.S. Introduction to Computational Biology: Maps, sequences and genomes. Chapman & Hall. London (1995).

Weeks et al., Plant Physiol., 102:1077 (1993).

Weissinger et al., Annual Rev. Genet., 22:421 (1988).

Wenzler et al., Plant Mol. Biol., 13:347 (1989).

White et al., Nucl Acids Res, 18:1062 (1990).

Yamamoto et al., Nucleic Acids Res., 18:7449 (1990).

Yu et al., Proc Natl Acad Sci USA 95:7819 (1998).

Zhang et al., Proc. Natl. Acad. Sci. USA, 94:4504 (1997).

All publications, patents and patent applications are incorporated herein by reference. While in the foregoing specification this invention has been described in relation to certain preferred embodiments thereof, and many details have been set forth for purposes of illustration, it will be apparent to those skilled in the art that the invention is susceptible to additional embodiments and that certain of the details described herein may be varied considerably without departing from the basic principles of the invention.

We claim:

1. An isolated nucleic acid molecule comprising a plant nucleotide sequence or its complement which hybridizes under low stringency conditions to a nucleic acid segment encoding a polypeptide comprising any one of SEQ ID NOs: 1-36, wherein the nucleotide sequence does not encode any one of SEQ ID NOs: 1-36.

2. An isolated nucleic acid molecule comprising a plant nucleotide sequence or its complement which hybridizes under high stringency conditions to a nucleic acid segment encoding a polypeptide comprising any one of SEQ ID NOs: 1-36, wherein the nucleotide sequence does not encode any one of SEQ ID NOs: 1-36.

3. An isolated nucleic acid molecule comprising a plant nucleotide sequence or its complement which hybridizes under moderate stringency conditions to a nucleic acid segment encoding a polypeptide comprising any one of SEQ ID NOs: 1-36, wherein the nucleotide sequence does not encode any one of SEQ ID NOs: 1-36.

4. An isolated nucleic acid molecule comprising a plant nucleotide sequence or its complement which encodes a polypeptide that is substantially similar to a polypeptide comprising any one of SEQ ID NOs: 1-36, wherein the nucleotide sequence does not encode any one of SEQ ID NOs: 1-36.

5. The isolated nucleic acid molecule of claim 1, 2, 3 or 4 which is DNA.

6. The isolated nucleic acid molecule of claim 1, 2, 3 or 4 which is RNA.

7. The isolated nucleic acid molecule of claim 4 wherein the nucleotide sequence encodes a polypeptide having at least 90% amino acid sequence identity to the polypeptide comprising any one of SEQ ID NOs: 1-36.

8. The isolated nucleic acid molecule of claim 4 wherein the nucleotide sequence encodes a polypeptide having at least 80% amino acid sequence identity to the polypeptide comprising any one of SEQ ID NOs: 1-36.

9. The isolated nucleic acid molecule of claim 4 wherein the nucleotide sequence encodes a polypeptide having at least 70% amino acid sequence identity to the polypeptide comprising any one of SEQ ID NOs: 1-36.

10. A polypeptide encoded by the nucleic acid molecule of claim 1, 2, 3 or 4.

11. An expression cassette comprising the nucleic acid molecule of claim 1, 2, 3 or 4 operably linked to suitable regulatory sequences.

12. The expression cassette of claim 11 which is linked to a promoter for expression in a plant.

13. A recombinant vector comprising the nucleic acid molecule of claim 1, 2, 3 or 4.

14. A host cell comprising the expression cassette of claim 11.

15. A host cell comprising the isolated nucleic acid molecule of claim 1, 2, 3 or 4.

16. The host cell of claim 15 which is selected from the group consisting of yeast, bacteria and plant.

17. A transformed plant, or seed thereof, the genome of which is augmented with the nucleic acid molecule of claim 1, 2, 3 or 4 which is expressed in an amount which confers increased protein content to the plant.

18. A transformed plant, or seed thereof, the genome of which is genetically altered so as to inhibit the expression of a gene corresponding to the nucleic acid molecule of claim 1, 2, 3 or 4.

19. The plan, or seed thereof, of claim 18 which is altered by T-DNA insertion, transposon insertion, or targeted DNA insertion.

20. The plant, or seed thereof, of claim 18 in which expression is inhibited by transcription or post-transcriptional mechanisms.

21. The plant, or seed thereof, of claim 17 or 18 which is a monocot.

22. The plant, or sed thereof, of claim 17 or 18 which is a dicot.

23. A method of expressing a nucleic acid molecule in a cell, comprising:
introducing the nucleic acid molecule of claim 1, 2, 3 or 4 into a cell so as to express the nucleic acid molecule.

24. The method of claim 23 wherein the cell is a plant cell.

25. The method of claim 23 wherein the cell is a monocot cell.

26. The method of claim 23 wherein the cell is a dicot cell.

27. A composition comprising the nucleic acid molecule of claim 1, 2, 3 or 4.

28. A composition comprising the polypeptide of claim 10.

29. A method to confer altered nutritional qualities to a plant, comprising:
a) contacting plant cells with an expression cassette comprising the nucleic acid molecule of claim 1, 2, 3 or 4 so as to yield transformed plant cells; and
b) regenerating the transformed plant cells to provide a differentiated transformed plant, wherein the differentiated transformed plant expresses the nucleic acid molecule in the cells of the plant in an amount effective to alter the protein content of the transformed plant relative to a corresponding plant which does not comprise the expression cassette.

30. A method to confer altered nutritional qualities to a plant, comprising:
a) contacting plant cells with an expression cassette comprising a nucleotide sequence encoding a polypeptide comprising any one of SEQ ID NOs: 1-36 so as to yield transformed plant cells; and

b) regenerating the transformed plant cells to provide a differentiated transformed plant, wherein the differentiated transformed plant expresses the nucleotide sequence in the cells of the plant in an amount effective to alter the protein content of the transformed plant relative to a corresponding plant which does not comprise the expression cassette.

31. A transformed plant prepared by the method of claim 29 or 30.

32. A seed of the plant of claim 31.

33. A progeny plant of the plant of claim 31.

34. An isolated nucleic acid molecule comprising a plant nucleotide sequence that directs transcription of an operatively linked nucleic acid fragment in a host cell, which nucleotide sequence corresponds to plant genomic DNA which is substantially similar to a nucleic acid segment which directs the transcription of a gene encoding a polypeptide comprising any one of SEQ ID NOs: 1-36.

35. The nucleic acid molecule of claim 34 wherein the nucleotide sequence has at least 90% identity to the nucleic acid segment.

36. A recombinant vector comprising the nucleic acid molecule of claim 34.

37. The vector of claim 36 which is a plasmid.

38. An expression cassette comprising the nucleic acid molecule of claim 34 operatively linked to an open reading frame.

39. The expression cassette of claim 38 operably linked to other suitable regulatory sequences.

40. A host cell comprising the expression cassette of claim 38.

41. A transformed plant, the genome of which is augmented with the expression cassette of claim 38.

42. A plant cell containing the expression cassette of claim 38.

43. A transformed plant comprising transformed plant cells, the transformed plant cells containing the expression cassette of claim 38.

44. The transformed plant of claim 43 wherein the plant is a dicot.

45. The cell of claim 42 which is a dicot cell.

46. The transformed plant of claim 43 wherein the plant is a monocot.

47. The cell of claim 42 which is a monocot cell.

48. The transformed plant of claim 43 which is a cereal plant.

49. A method of augmenting a plant genome, comprising:

a) contacting plant cells with the expression cassette of claim 38 so as to yield transformed plant cells; and

b) regenerating the transformed plant cells to provide a differentiated transformed plant, wherein the differentiated transformed plant expresses the nucleic molecule in the cells of the plant.

50. A transformed plant prepared by the method of claim 49.

51. A seed of the plant of claim 50.

52. A progeny plant of the plant of claim 50.

53. A method of using a plant promoter, comprising: introducing the expression cassette of claim 38 to a plant cell and detecting the expression of the product of the open reading frame.

54. An isolated nucleic acid molecule comprising a plant nucleotide sequence that directs transcription of an operatively linked nucleic acid fragment in a plant cell, which nucleotide sequence corresponds to plant genomic DNA which hybridizes under low stringency conditions a nucleic acid segment that directs transcription of a gene encoding a polypeptide comprising any one of SEQ ID NOs: 1-36.

55. An isolated nucleic acid molecule comprising a plant nucleotide sequence that directs transcription of an operatively linked nucleic acid fragment in a plant cell, which nucleotide sequence corresponds to plant genomic DNA which hybridizes under high stringency conditions a nucleic acid segment that directs transcription of a gene encoding a polypeptide comprising any one of SEQ ID NOs: 1-36.

56. A recombinant vector comprising the expression cassette of claim 38.

57. A plant cell comprising the vector of claim 56.

58. A transformed plant, the cells of which comprise the vector of claim 56.

59. The nucleic acid molecule of claim 34, 54 or 55 wherein the nucleotide sequence is 25 to 2000 nucleotides in length.

60. The expression cassette of claim 38 wherein the open reading frame is in an antisense orientation.

61. The expression cassette of claim 38 wherein the open reading frame is in a sense orientation.

62. The expression cassette of claim 12 wherein the nucleic acid molecule is in antisense orientation.

63. The expression cassette of claim 12 wherein the nucleic acid molecule is in sense orientation.

64. An antibody that binds to the polypeptide of claim 10.

65. A method for marker-assisted selection of plants having a desired property, comprising:

a) contacting a probe comprising at least a portion of a nucleic acid sequence comprising an open reading frame encoding a polypeptide comprising any one of SEQ ID NOs: 1-36 with a nucleic acid sample from a plant in an amount sufficient to form complexes; and

b) detecting or determining the amount of complex formation.

66. A method for marker-assisted selection of plants having a desired property, comprising:

a) contacting a probe comprising at least a portion of the nucleic acid molecule of claim 1, 2, 3 or 4 with a nucleic acid sample from a plant in an amount sufficient to form complexes; and

b) detecting or determining the amount of complex formation

67. A method for marker-assisted selection of plants having a desired property, comprising:

a) contacting a sample comprising plant proteins with the antibody of claim 64 in an amount sufficient to form complexes; and

b) detecting or determining the amount of complex formation.

68. A method to identify transcription factors for genes associated with high protein content in plants, comprising:

a) contacting the nucleic acid molecule of claim 34, 54 or 55 with a sample comprising transcription factor polypeptides so as to form a complex between the nucleic acid molecule and at least one transcription factor; and

b) detecting or determining complex formation.

69. The method of claim 68 further comprising identifying the transcription factor in the complex.

70. A method of feeding livestock, which comprises feeding livestock a plant of any of the claims 17-22, 31-33, 41-48, 50-52 or 57-58, or a plant part thereof.

71. The method of claim 70 where the plant part is grain or seed.

72. A manufacturing process, which comprises milling grain produced on a plant of any of the claims 17-22, 31-33, 41-48, 50-52 or 57-58, or a plant part thereof.

73. A product which comprises milled grain produced on a plant of any of the claims 17-22, 31-33, 41-48, 50-52 or 57-58, or a plant part therof.

74. The product of claim 73 that is a human or animal food product.

75. A method of producing an industrially or therapeutically important protein in a plant or part thereof, such as a seed, comprising modulaton of high protein phenotype by over or under expressing one or more high protein phenotype genes in the host plant.

76. The method of claim 78 wherein under expressing or down regulation of one or more of the genes provides a plant or part thereof, with an increased ability to produce the industrially or therapeutically important polypeptide.

Figure 1

Seed protein content in the high protein lines and in selected hybrids

Maize materials	Protein %
1 Inbreds	
LH59	13.5
WI500	15.6
2. Hybrids	
73/93	11.9
73/88	11.9
73/98	11.9
73/92	8.2
73/76	8.4
73/84	8.5

The high protein materials used in this study were from Wilson Genetics, originally derived from a tropical germplasm. WIL500 is the key high protein source and was the inbred used in most of this study for high protein source. The normal protein control used in this study is LH59. Seeds of LH59 and WIL 500 were generated from green house generated seed materials

The hybrids listed in this study were derived from recombinant inbred lines crossed with a tester line called JHAx412B. The inbred lines used to generate the hybrids were recombinant inbred lines (F4 generation) generated from a cross between WI500 and a normal protein line, BIIJ208.

FIGURE 2A

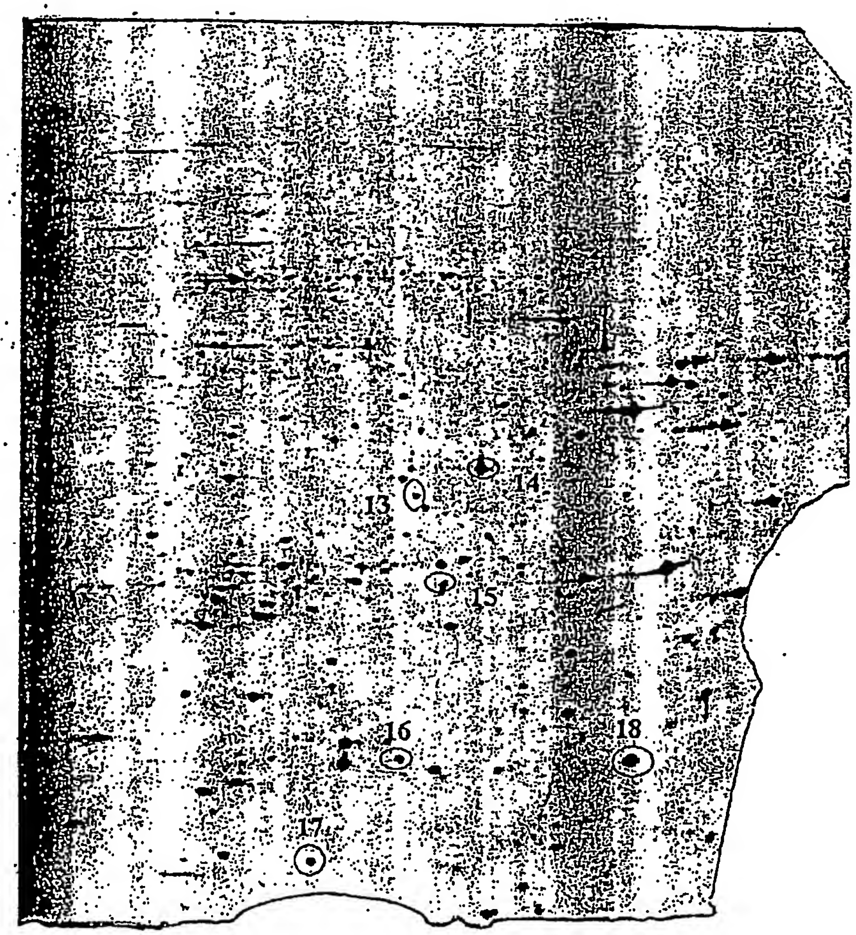


FIGURE 2B

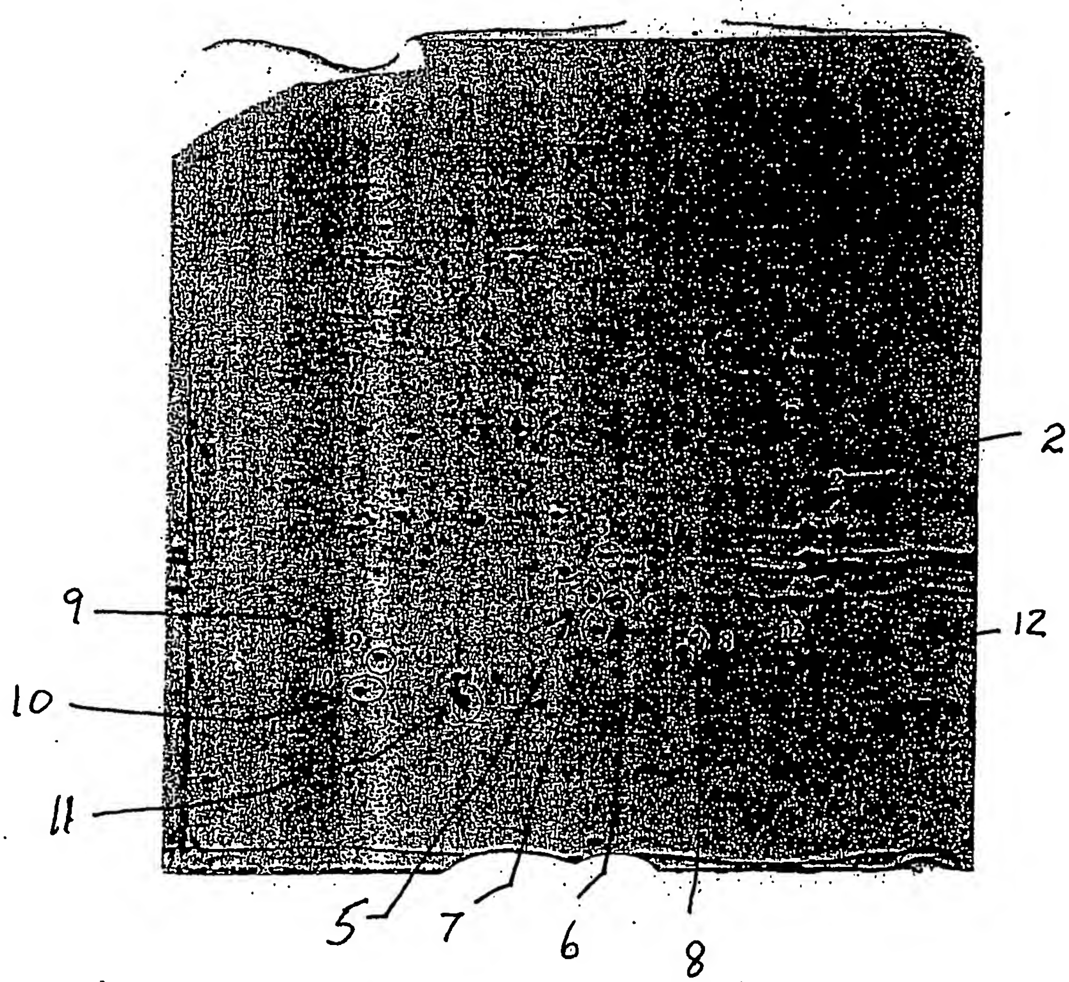
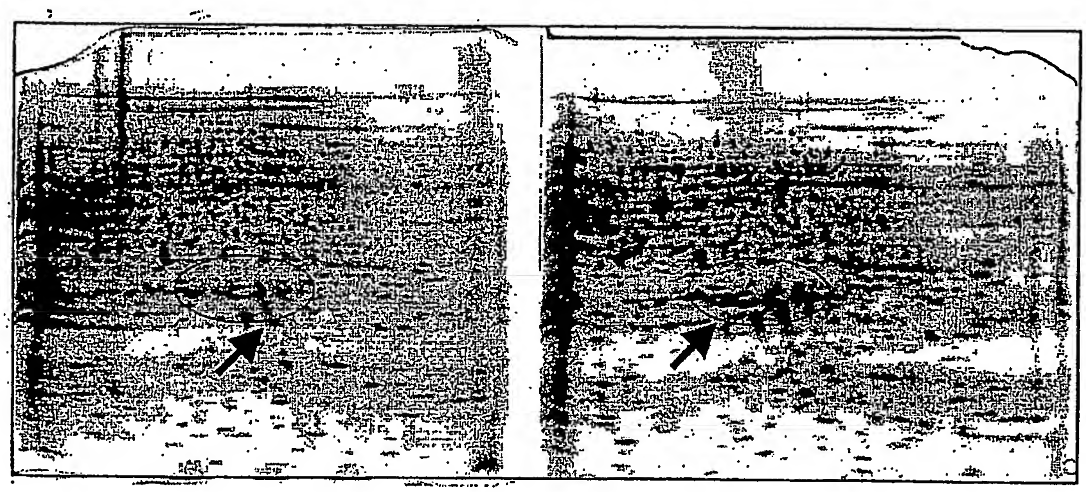


Figure 2C and 2D

Comparision of protein expression profile of high protein germplasm and normal corn line

This example demonstrated the proteomic approach used to identify proteins that are related to the high protein phenotype. Seed proteins were separated by two dimensional electrophoresis and the differentially expressed spots were identified by mass spec as described in the methods. 30DAP embryos were used. The arrow points to a clear difference area that contains the various forms of globulins proteins in embryo.

2C 2D



LH 59 WI500

FIGURE 3A

Spot wp0824a_1 **Corn seed embryo**

Apparent mass: 50.0, apparent pI: 6.5

XCorr	dCn	Sp	RSp	Ions	Ref	Sequence1664.8 (-0.5)
4.4007	0.425	2421.4	1	21/26	PIR2:S15675	R.FTHELLEDAVGNYR.V

PIR2:S15675 globulin-2 precursor - maize

MKVPVLLLLV SLCFSLALAW QTDTESGSGR PYHYGEESFR HWTRSRQGRF
RVLERFTHEL LEDAVGNYRV AELEAAPRAF LQPSHYDADE VMFVKEGEGV
IVLLRGGKRE SFCVREGDVM VIPAGAVVYS ANTHQSEWFR VVMLLSPVVS
TSGRFEEFFP IGGESPESFL SVFSDDVIQA SFNTRREEWE KVFEKQSKGE
ITTASEEQIR ELSRSCSRGG RSSRSEGGDS GSSSSKWEIK PSSLTDKKPT
HSNSHGRHYE ITGDECPHLR LLDMDVGLAN IARGSMMAPS YNTRANKIAI
VLKGQGYFEM ACPHVSGGRS SPRRERGHGR EEEEEREEEQ GGGGGQKSRS
YRQVKSRIRE GSVIVIPAGH PTALVAGEDK NLAVLCFEVN ASFDDKVFLA
GTNSALQKMD RPAKLLAFGA DEEQQVDRVI GAQKDAVFLR GPQSHRVSSV

Position	MH+	Sequence (link:NCBI Blast)
56-69	1664.8149	FTHELLEDAVGNYR

Spot wp0824a_4 **corn seed embryo**

Apparent mass: 17.9 apparent pI: 5.8
3.4360 0.365 1424.1 2 19/26 PIR2:S72545 +7 N.AGLENGVLTVTVPK.A

PIR2:S72545 heat shock protein 16.9 - pearl millet

MSLVRRGNVF DPFSMDLWDP FDNMFRSIVP SSSSSDTAAF ANARIDWKET
PEVHVFKADL PGVKKEEVKV EVEDGNVLVI SGQRSKEKED KNDRWHRVER
SSGQFVRRFR LPEDAKTDQV NAGLENGVLT VTVPKAEGKK PEVKAIEISG

Position	MH+	Sequence (link:NCBI Blast)
122-13	1398.6426	AGLENGVLTVTVPK

FIGURE 3B

Spot wp0824a_6 corn seed embryo

Apparent mass: 21.5 apparent pI: 6.2
3.6105 0.302 1575.8 1 18/26 PIR2:T04358 K.GLAYEYLEQDLGNK.S

PIR2:T04358 glutathione transferase (EC 2.5.1.18)

MAEEKKQGLQ LLDFWVSPFG QRCRIAMDEK **GLAYEYLEQD LGNK**SELLLR
ANPVHKKIPV LLHDGRPVCE SLVIVQYLDE AFPAAAPALL PADPYARAQA
RFWADYVDKK LYDCGTRLWK LKGDGQAQAR AEMVEILRTL EGALGDGPFF
GGDALGFVDV ALVPFTSWFL AYDRFGGVSV EKECPRLAAW AKRCAERPSV
AKNLYPPEKV YDFVCGMKKR LGIE

Position	MH+	Sequence (link:NCBI Blast)
31-44	1613.7642	GLAYEYLEQDLGNK

Spot wp0810w_11 corn seed embryo

Apparent mass: 26.0 apparent pI: 8.4

3.7625 0.367 1493.8 1 19/30 X14312.1_0 +1 K.IDLQTAQQLQNQDDNR.G

X14312.1_0 Arabidopsis CRA1 gene for 12S seed storage protein. //start:stop=196:1951
//PID=; Arabidopsis
thaliana; 12S seed storage protein
MARVSSLLSF CLTLLILFHG YAAQQGQQGQ QFPNECQLDQ LNALEPSHVL
KSEAGRIEVW DHHAPQLRCS GVSFARYIIE SKGLYLPSFF NTAKLSFVAK
GRGLMGKVIP GCAETFQDSS EFQPRFEGQG QSQRFRDMHQ KVEHIRSGDT
IATTPGVAQW FYNDGQQPLV IVSVFDLASH QNQLDRNPRP FYLAGNNPQG
QVWLQGREQQ PQKNIFNGFG PEVIAQALKI **DLQTAQQLQN QDDNR**GNIVR
VQGPFGVIRP PLRGQRPQEE EEEEGRHGRH GNGLEETICS ARCTDNLDDP
SRADVYKPQL GYISTLNSYD LPILRFIRLS ALRGSIRQNA MVLPQWNANA
NAILYETDGE AQIQIVNDNG NRVFDGQVSQ GQLIAVPQGF SVVKRATSNR
FQWVEFKTNA NAQINTLAGR TSVLRGLPLE VITNGFQISP EEARRVKFNT
LETTLTHSSG PASYGRPRVA AA

Position	MH+	Sequence (link:NCBI Blast)
230-245	1900.9998	IDLQTAQQLQNQDDNR

FIGURE 3C

Spot wp0801w_21 corn seed embryo

Apparent mass: 75.0 apparent pI: 5.5

3.1818 0.377 1116.3 1 20/ 42 AF121355.1_0 +1
K.VTVANVESGGEFTVSSADDILK.A

AF121355.1_0 Arabidopsis thaliana peroxiredoxin TPx1 mRNA, complete cds.
//start:stop=54:542 //PID=; Arabidopsis
thaliana; thioredoxin-dependent peroxidase; peroxiredoxin TPx

MAPIAVGDVV PDGTISFFDE NDQLQTASVH SLAAGKKVIL FGVPGAFTPT
CSMKHVPGFI EKAEELKSKG VDEIICFSVN DPFVMKAWGK TYPENKHVKF
VADGSGEYTH LLGLELDLKD KGLGVRSRRF ALLLDDLKVT **VANVESGGEF
TVSSADDILK** AL

Position	MH+	Sequence (link:NCBI Blast)
139-160	2239.4414	VTVANVESGGEFTVSSADDILK

FIGURE 3D

Spot nla0808_11 corn seed embryo

Apparent mass: 65.0 apparent pI: 6.6
:
3.4352 0.173 2003.0 1 18/ 20 SW:GLB1_MAIZE +3 K.AEEVDEVLGSR.R

SW:GLB1_MAIZE P15590 zea mays (maize). globulin-1 s allele precursor (glb1-s) (7s-like). 7/1999
MVSARIVVLL AVLLCAAAAV ASSWEDDNHH HHGGHKSGRC VRRCEDRPWH
QRPRCLEQCR EEEREKRQER SRHEADDRSG EGSSEDERER EQEKEEKQKD
RRPYVFDRRS FRRVVRSEQG SLRVLRPFDE VSRLLRGIRD YRVAVLEANP
RSFVVPSHTD AHCIGYVAEG EGVVTTIENG ERRSYTIKQG HVFVAPAGAV
TYLANTDGRK KLVITKILHT ISVPGEFQFF FGPGGRNPES FLSSFSKSIQ
RAAYKTSSDR LERLFGRHGQ DKGIIVRATE EQTRELRRHA SEGGHGPHWP
LPPFGESRGP YSLLDQRPSI ANQHGQLYEA DARSFHDLAE HDVSVSFANI
TAGSMSAPLY NTRSFKIAYV PNGKGYAEIV CPHRQSQGGE SERERGKGRR
SEEEEESSEE QEEVGQGYHT IRARLSPGTA FVVPAGHPFV AVASRDSNLQ
IVCFEVHADR NEKVFLAGAD NVLQKLDRVA KALSFASKAE **EVDEVLGSR**R
EKGFLPGPKE SGGHEEREQE EEEREERHGG RGERERHGRE EREKEEEERE
GRHGRGRREE VAETLLRMVT ARM

Position	MH+	Sequence (link:NCBI Blast)
489-499	1204.2793	AEEVDEVLGSR

FIGURE 3E

Spot nla0808_4 corn seed embryo

Apparent mass: 65.0 apparent pI:.6.8

4.3004 0.354 2163.0 1 36/72 PIR2:S18545
R.NDGTARPGGVAASMAAAAR.L

PIR2:S18545 rab28 protein - maize

MSQEQPRRPS GHEETSGGGE QGAVRYGDVF PAVSGGLAEK PVARRTATMQ
SAENLVFGQT LKGGPAAAMQ SAATTNERMG AVGHDQATDA TAVQGVTVSE
TRVPGGGRIV TEFVAGQAVG QYLARDDDGG GGIAGPGAGA GVAGKDITKV
TIGEALEATA LAAGDAPVER SDAARIQAAE ARATGLDANV PGGLARQAQS
AAAANSWAWG DEDKATLGDV LANATARLVA DKPVESADAL GVAGAENRNR
NDGTARPGGV AASMAAAARL NRNEAVWE

Position	MH+	Sequence (link:NCBI Blast)
251-269	1744.9233	NDGTARPGGVAASMAAAAR

FIGURE 3F

Spot nla0808_7 **corn seed embryo**

Apparent mass: 55.0 apparent pI: 6.8

4.4214 0.477 1987.8 1 24/ 36 SW:ZEAD_MAIZE +1
R.AQQLQQLVLANLAAYSQQH.Q
SW:ZEAD_MAIZE P24450 zea mays (maize). zein-alpha precursor (19 kd) (pms2).

MAAKIFCFLM LLGLSASVAT ATIFPQCSQA PIASLLPPYL SPAVSSMCEN PIVQPYRIQQ AIATGILPLS PLFLQQPSAL LQQLPLVHLV AQNIRA**QQLQ QLVLANLAAY SQQH**QFLPFN QLAALNSAAY LQQQLPFSQL VAAYPRQFLP FNQLAALNSA AYLQQQQLLP FSQLADVSPA AFLTQQQLLP FYLHAMPNAG TLLQLQQLLP FNQLALTNST VFYQQPIIGG ALF

3.2085 0.226 2050.3 1 17/ 20 SW:ZEAA_MAIZE R.LQQAIAASILR.S

SW:ZEAA_MAIZE P06678 zea mays (maize). zein-alpha precursor (19 kd) (clone 19d1). 6/1994
MAAKIFALLA LLALSANVAT ATIIPQCSQQ YLSPVTAARF EYPTIQSYRL **QQAIAASILR** SLALTVQQPY ALLQQPSLVN LYLQRIVAQQ LQQQLLPTIN QVVAANLDAY LQQQQFLPFN QLAGVNPAAY LQAQQLLPFN QLVRSPAAFL LQQQLLPFHL QVVANIAAFL QQQQLLPFYP QVVGNINAFL QQQQLLPFYP QDVANNVAFL QQQQLLPFSQ LALTNPTTLL QQPTIGGAIF

Position	MH+	Sequence (link:NCBI Blast)
96-114	2125.3921	AQQLQQLVLANLAAYSQQH
50-60	1184.4243	LQQAIAASILR

Spot nla0808_8 **corn seed embryo**

Apparent mass: 21.0 apparent pI: 6.9

2.7962 0.324 1236.8 1 17/ 26 SW:OLE1_MAIZE R.GATGGGGGYGDLQR.G

SW:OLE1_MAIZE P13436 zea mays (maize). oleosin zm-i (oleosin 16 kd) (lipid body-associated major protein) (lipid
body-associated protein l3). 7/1998

MADHHR**GATG GGGGYGDLQR** GGGMHGEAQQ QQKQGAMMTA LKAATAATFG GSMLVLSGLI LAGTVIALTV ATPVLVIFSP VLVPAAIALA LMAAGFVTSG
GLGVAALSVF SWMYKYLTGK HPPAADQLDH AKARLASKAR DVKDAAQHRI DQAQGS

Position	MH+	Sequence (link:NCBI Blast)
7-20	1266.3126	GATGGGGGYGDLQR

FIGURE 3G

Spot nla0907_1 corn seed embryo

Apparent mass: 60.0 apparent pI: 6.0

2.9706 0.127* 954.7 1 17/ 24 PIR2:T06212 K.VALVTGGDSGIGR.A

PIR2:T06212 glucose and ribitol dehydrogenase homolog - barley
MASQKFPPQQ QDCQPGKEHA MDPRPEAIIK NYKSANKLQG **KVALVTGGDS**
GIGRAVCLCL ALEGATVNFT YVKGHEDKDA EETLQALRDI KSRTGAGEPK
ALSGDLGYEE NCRRVVEEVA NAHGGRVDIL VNNAAEQYVR PCITEITEQD
LERVFRTNIF SYFLMTKFAV KHMGPGSSII NTTSVNAYKG NATLLDYTAT
KGAIVAFTRA LSMQLAEKGI RVNGVAPGPI WTPLIPASFP EEKVKQFGSE
VPMKRAGQPS EVAPSFVFLA SEQDSSYISG QILHPNGGTI VNS

Position	MH+	Sequence (link:NCBI Blast)
42-54	1202.353	VALVTGGDSGIGR

Spot nla0907_12 corn seed embryo

Apparent mass: 16.5 apparent pI: 7.0

3.7833 0.110 1211.9 1 27/ 48 SW:HS11_MEDSA +8 R.IDWKETPEAHVFK.A

SW:HS11_MEDSA P27879 medicago sativa (alfalfa). 18.1 kd class i heat shock protein (fragment). 4/1993
DPFSLDVWDP FKDFPFTNSA LSASSFPQEN SAFVSTR**IDW KETPEAHVFK**
ADLPGLKKEE VKVEIEDDRV LQISGERNVE KEDKNDQWHR VERSSGKFMR
R**FRLPENAK**M DQVKAAMENG VLTVTVPKEE IKKPEVKSIE ISS

Position	MH+	Sequence (link:NCBI Blast)
102-109	975.1355	FRLPENAK
38-50	1600.8143	IDWKETPEAHVFK

FIGURE 3H

Spot nla0907_16 corn seed embryo

Apparent mass: 21.5 apparent pI: 6.2

3.2768 0.167 1376.0 1 17/26 PIR2:T04358 K.GLAYEYLEQDLGNK.S

PIR2:T04358 glutathione transferase (EC 2.5.1.18) - maize
MAEEKKQGLQ LLDFWVSPFG QRCRIAMDEK **GLAYEYLEQD LGNK**SELLLR
ANPVHKKIPV LLHDGRPVCE SLVIVQYLDE AFPAAAPALL PADPYARAQA
RFWADYVDKK LYDCGTRLWK LKGDGQAQAR AEMVEILRTL EGALGDGPFF
GGDALGFVDV ALVPFTSWFL AYDRFGGVSV EKECPRLAAW AKRCAERPSV
AKNLYPPEKV YDFVCGMKKR LGIE

Position	MH+	Sequence (link:NCBI Blast)
31-44	1613.7642	GLAYEYLEQDLGNK

Spot nla0907_7 corn seed embryo

Apparent mass: 25.0 apparent pI: 6.6

5.4645 0.423 2528.5 1 39/100 PIR2:S05545
H.GHGATGHVDQYGNPVGGVEHGTGGMR.H

PIR2:S05545 dehydrin 3 - maize
MEYGQQGQHG **HGATGHVDQY GNPVGGVEHG TGGMR**HGTGT
GGMGQLGEHG GAGMGGGQFQ PAREEHKTGG ILHRSGSSSS SSSEDDGMGG
RRKKGIKEKIKEKLPGGHKD DQHATATTGG AYGQQGHTGS AYGQQGHTGG
AYATGTEGTG EKKGIMDKIK EKLPGQH

Position	MH+	Sequence (link:NCBI Blast)
10-35	2548.7119	GHGATGHVDQYGNPVGGVEHGTGGMR

Figure 4A

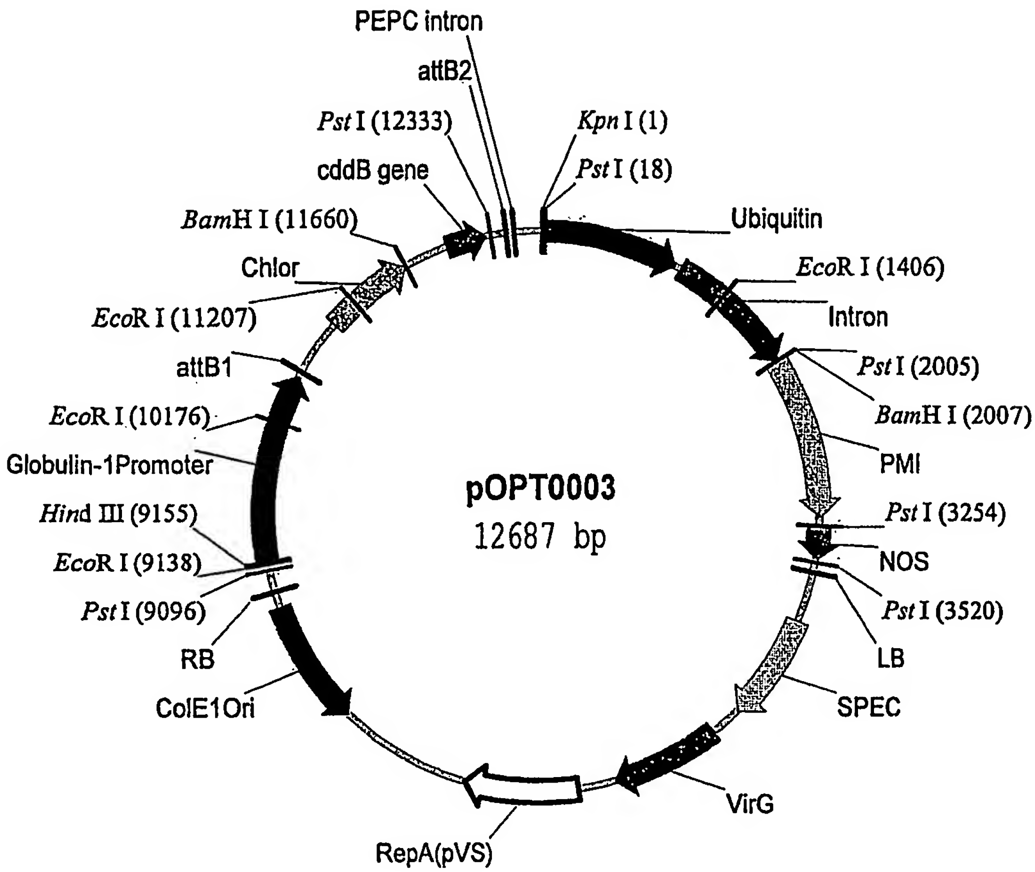

PEPC intron
attB2
PstI (12333)
KpnI (1)
PstI (18)
cddB gene
BamHI (11660)
Ubiquitin
EcoRI (1406)
Chlor
EcoRI (11207)
Intron
attB1
PstI (2005)
BamHI (2007)
EcoRI (10176)
Globulin-1Promoter
PMI
pOPT0003
12687 bp
HindIII (9155)
PstI (3254)
EcoRI (9138)
NOS
PstI (9096)
PstI (3520)
RB
LB
ColE1Ori
SPEC
VirG
RepA(pVS)

Figure 4B

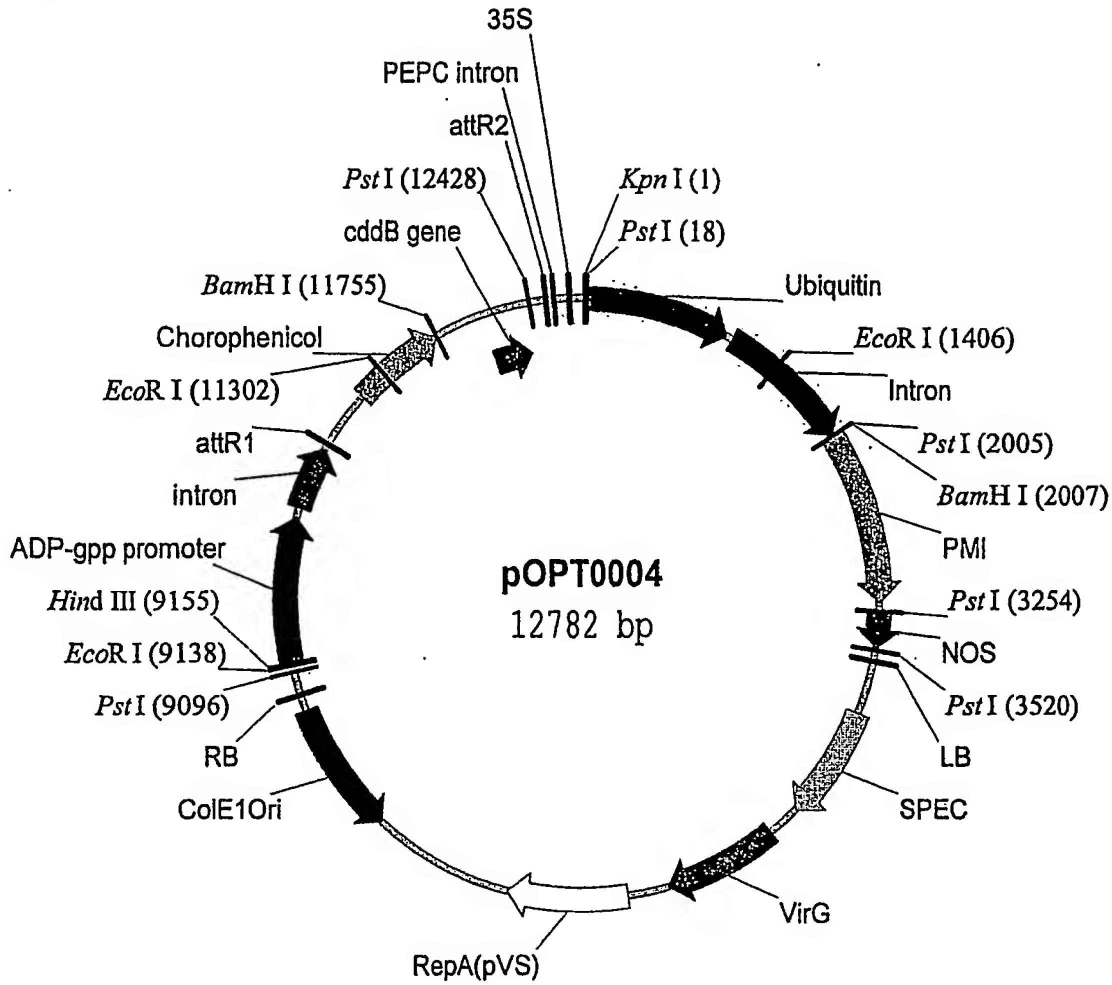
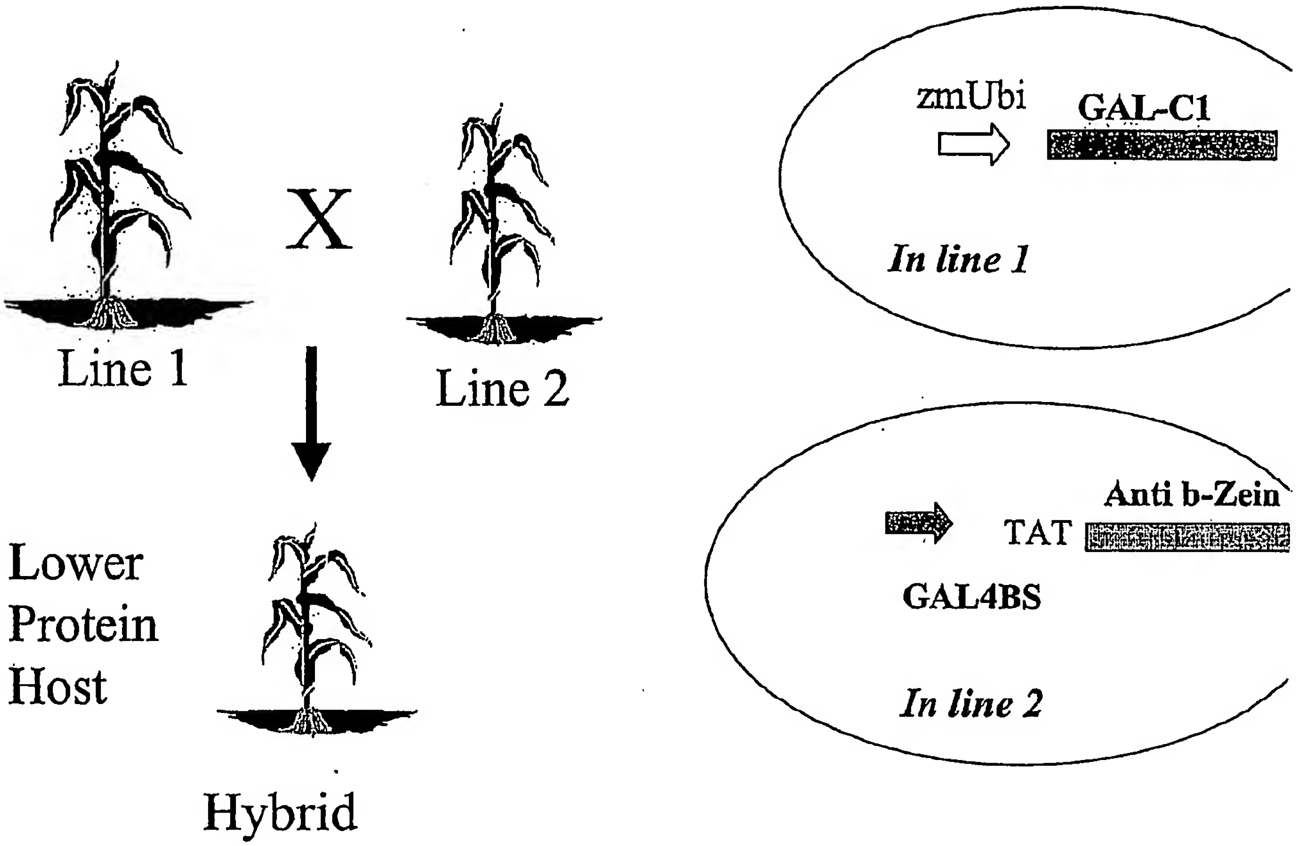

35S
PEPC intron
attR2
PstI (12428)
KpnI (1)
PstI (18)
cddB gene
BamH I (11755)
Ubiquitin
Chorophenicol
EcoR I (1406)
EcoR I (11302)
Intron
attR1
PstI (2005)
intron
BamH I (2007)
ADP-gpp promoter
PMI
pOPT0004
12782 bp
Hind III (9155)
PstI (3254)
EcoR I (9138)
NOS
PstI (9096)
PstI (3520)
RB
LB
ColE1Ori
SPEC
VirG
RepA(pVS)

Figure 5
Producing lower storage protein maize grain



Sequence Listing

<110> Su, Wenpei
Andon, Nancy
Haynes, Paul
Briggs, Steven P.
Cooper, Bret
Glazebrook, Jane
Goff, Stephen A.
Katagiri, Fumiaki
Kreps, Joel
Moughamer, Todd
Provart, Nicholas
Ricke, Darrell
Zhu, Tong

<120> High-protein-phenotype-associated plant genes

<130> 60011-PCT

<140>

<141> 2002-09-26

<150> US 60/325,277
<151> 2001-09-26

<150> US 60/370,526
<151> 2002-04-04

<150> US 60/370,620
<151> 2002-04-04

<160> 45

<170> PatentList.pl version 3.0.4 (C) 2001 Syngenta

<210> 1
<211> 1722
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (1)..(1722)
<223> GLOBULIN-1 S ALLELE PRECURSOR
gene:GLB1-S
sp|P15590|GLB1_MAIZE

<400> 1
atg gtg agc gcc aga atc gtt gtc ctc ctc gcc gtc ctc cta tgc gct 48
Met Val Ser Ala Arg Ile Val Val Leu Leu Ala Val Leu Leu Cys Ala
1 5 10 15

gcc gcc gca gtc gcg tcg tcc tgg gag gac gac aac cac cac cac cac 96
Ala Ala Ala Val Ala Ser Ser Trp Glu Asp Asp Asn His His His His
20 25 30

ggg ggc cac aag tcc ggg cga tgc gtg cgg cgg tgc gag gac cgg ccc 144
Gly Gly His Lys Ser Gly Arg Cys Val Arg Arg Cys Glu Asp Arg Pro
35 40 45

```
tgg cac cag cgc ccc cgg tgc ctg gag cag tgc agg gag gag gag cgg      192
Trp His Gln Arg Pro Arg Cys Leu Glu Gln Cys Arg Glu Glu Glu Arg
    50                  55                  60

gag aag cgg caa gag cgg agc agg cac gag gcc gac gac cgc agc ggc      240
Glu Lys Arg Gln Glu Arg Ser Arg His Glu Ala Asp Asp Arg Ser Gly
65                  70                  75                  80

gag ggc tcg tcg gag gat gag cgc gag cgc gag cag gag aag gag gag      288
Glu Gly Ser Ser Glu Asp Glu Arg Glu Arg Glu Gln Glu Lys Glu Glu
                85                  90                  95

aag cag aag gac cgg cgg ccg tac gtg ttc gac cgg cgc agc ttt cgt      336
Lys Gln Lys Asp Arg Arg Pro Tyr Val Phe Asp Arg Arg Ser Phe Arg
            100                 105                 110

cgc gtg gtc cgg agc gag cag ggg tcc ctg agg gtg ctc cgg ccg ttc      384
Arg Val Val Arg Ser Glu Gln Gly Ser Leu Arg Val Leu Arg Pro Phe
        115                 120                 125

gac gag gtg tcc agg ctc ctc cgc ggc atc cgg gac tac cgc gtg gcg      432
Asp Glu Val Ser Arg Leu Leu Arg Gly Ile Arg Asp Tyr Arg Val Ala
    130                 135                 140

gtc ctg gag gcg aac ccg cgc tcg ttc gtg gtg ccc agc cac acc gac      480
Val Leu Glu Ala Asn Pro Arg Ser Phe Val Val Pro Ser His Thr Asp
145                 150                 155                 160

gcg cac tgc atc ggc tac gtg gcg gaa ggc gag gga gtg gtg acg acg      528
Ala His Cys Ile Gly Tyr Val Ala Glu Gly Glu Gly Val Val Thr Thr
                165                 170                 175

atc gag aac ggc gag agg cgg tcg tac acc atc aag caa ggc cac gtc      576
Ile Glu Asn Gly Glu Arg Arg Ser Tyr Thr Ile Lys Gln Gly His Val
            180                 185                 190

ttc gtg gcg ccg gcc ggg gcg gtc acc tac ctg gcc aac acc gac ggc      624
Phe Val Ala Pro Ala Gly Ala Val Thr Tyr Leu Ala Asn Thr Asp Gly
        195                 200                 205

cgg aag aaa ctg gtc atc acc aag atc ctc cat acc atc tcc gtg cct      672
Arg Lys Lys Leu Val Ile Thr Lys Ile Leu His Thr Ile Ser Val Pro
    210                 215                 220

ggc gag ttc cag ttc ttc ttc ggc ccc ggc ggg agg aac ccg gaa tcg      720
Gly Glu Phe Gln Phe Phe Phe Gly Pro Gly Gly Arg Asn Pro Glu Ser
225                 230                 235                 240

ttc ctg tcg agc ttc agc aag agc atc cag aga gct gcg tac aag acc      768
Phe Leu Ser Ser Phe Ser Lys Ser Ile Gln Arg Ala Ala Tyr Lys Thr
                245                 250                 255

tcg agc gac cgg ctg gag agg ctg ttc ggg agg cat ggg cag gac aag      816
Ser Ser Asp Arg Leu Glu Arg Leu Phe Gly Arg His Gly Gln Asp Lys
            260                 265                 270

ggg atc atc gtg cgt gcc acg gag gag cag acc cgc gag ctg cgg cgc      864
Gly Ile Ile Val Arg Ala Thr Glu Glu Gln Thr Arg Glu Leu Arg Arg
        275                 280                 285
```

```
cac gcc tcg gag ggc ggc cac ggc ccg cac tgg ccc ctg ccg ccg ttc      912
His Ala Ser Glu Gly Gly His Gly Pro His Trp Pro Leu Pro Pro Phe
    290                 295                 300

ggc gag tcg cgc ggc ccc tac agc ctc ctg gac cag cgg ccc agc atc      960
Gly Glu Ser Arg Gly Pro Tyr Ser Leu Leu Asp Gln Arg Pro Ser Ile
305                 310                 315                 320

gcc aac cag cac ggg cag ctc tac gag gcc gac gcg cgc agc ttc cac     1008
Ala Asn Gln His Gly Gln Leu Tyr Glu Ala Asp Ala Arg Ser Phe His
                325                 330                 335

gac ctc gcc gag cac gac gtc agc gtc tcc ttc gcc aac atc acc gcg     1056
Asp Leu Ala Glu His Asp Val Ser Val Ser Phe Ala Asn Ile Thr Ala
            340                 345                 350

ggg tcc atg agc gcg cca ttg tac aac acc cgt tcg ttc aag atc gcc     1104
Gly Ser Met Ser Ala Pro Leu Tyr Asn Thr Arg Ser Phe Lys Ile Ala
        355                 360                 365

tac gtg ccg aac ggc aag ggc tac gcc gag atc gtg tgc ccg cac cgc     1152
Tyr Val Pro Asn Gly Lys Gly Tyr Ala Glu Ile Val Cys Pro His Arg
    370                 375                 380

cag tcg cag ggc ggc gag agc gag cgc gag cgc ggc aag ggc agg agg     1200
Gln Ser Gln Gly Gly Glu Ser Glu Arg Glu Arg Gly Lys Gly Arg Arg
385                 390                 395                 400

agc gaa gaa gaa gaa gaa tcg tct gag gag cag gag gaa gtc ggg cag     1248
Ser Glu Glu Glu Glu Glu Ser Ser Glu Glu Gln Glu Glu Val Gly Gln
                405                 410                 415

ggg tac cac acc atc cgg gcg cgg ctg tca ccg ggc acg gcg ttc gtg     1296
Gly Tyr His Thr Ile Arg Ala Arg Leu Ser Pro Gly Thr Ala Phe Val
            420                 425                 430

gtg ccc gcg ggc cac ccg ttc gtc gcg gtg gcg tcc cgg gac agc aac     1344
Val Pro Ala Gly His Pro Phe Val Ala Val Ala Ser Arg Asp Ser Asn
        435                 440                 445

ctc cag atc gtg tgc ttc gag gtc cac gcc gac agg aac gag aag gtg     1392
Leu Gln Ile Val Cys Phe Glu Val His Ala Asp Arg Asn Glu Lys Val
    450                 455                 460

ttc ctg gcc ggc gcc gac aac gtg ctg cag aag ctc gac cgg gtc gcc     1440
Phe Leu Ala Gly Ala Asp Asn Val Leu Gln Lys Leu Asp Arg Val Ala
465                 470                 475                 480

aag gcg ctg tca ttc gcc tcc aag gcg gag gag gtg gac gag gtg ctc     1488
Lys Ala Leu Ser Phe Ala Ser Lys Ala Glu Glu Val Asp Glu Val Leu
                485                 490                 495

ggc tcg cgg cgc gag aag ggg ttc ctt cct ggc ccc aag gag agc ggc     1536
Gly Ser Arg Arg Glu Lys Gly Phe Leu Pro Gly Pro Lys Glu Ser Gly
            500                 505                 510

ggc cac gag gag cgg gag cag gag gag gag gaa cgc gaa gaa cgc cac     1584
Gly His Glu Glu Arg Glu Gln Glu Glu Glu Glu Arg Glu Glu Arg His
        515                 520                 525

ggc ggg cgt ggg gag agg gaa cgc cac gga cgt gag gag cgg gag aaa     1632
```

Gly Gly Arg Gly Glu Arg Glu Arg His Gly Arg Glu Glu Arg Glu Lys
530 535 540

gag gag gag gaa cgc gaa gga cgc cac ggc cgc ggg cgc cgc gag gaa 1680
Glu Glu Glu Glu Arg Glu Gly Arg His Gly Arg Gly Arg Arg Glu Glu
545 550 555 560

gtg gcg gag acg ctc ctg agg atg gtg acc gcc agg atg tga 1722
Val Ala Glu Thr Leu Leu Arg Met Val Thr Ala Arg Met
565 570

<210> 2
<211> 573
<212> PRT
<213> Zea mays

<400> 2

Met Val Ser Ala Arg Ile Val Val Leu Leu Ala Val Leu Leu Cys Ala
1 5 10 15

Ala Ala Ala Val Ala Ser Ser Trp Glu Asp Asp Asn His His His His
20 25 30

Gly Gly His Lys Ser Gly Arg Cys Val Arg Arg Cys Glu Asp Arg Pro
35 40 45

Trp His Gln Arg Pro Arg Cys Leu Glu Gln Cys Arg Glu Glu Glu Arg
50 55 60

Glu Lys Arg Gln Glu Arg Ser Arg His Glu Ala Asp Asp Arg Ser Gly
65 70 75 80

Glu Gly Ser Ser Glu Asp Glu Arg Glu Arg Glu Gln Glu Lys Glu Glu
85 90 95

Lys Gln Lys Asp Arg Arg Pro Tyr Val Phe Asp Arg Arg Ser Phe Arg
100 105 110

Arg Val Val Arg Ser Glu Gln Gly Ser Leu Arg Val Leu Arg Pro Phe
115 120 125

Asp Glu Val Ser Arg Leu Leu Arg Gly Ile Arg Asp Tyr Arg Val Ala
130 135 140

Val Leu Glu Ala Asn Pro Arg Ser Phe Val Val Pro Ser His Thr Asp
145 150 155 160

Ala His Cys Ile Gly Tyr Val Ala Glu Gly Glu Gly Val Val Thr Thr
165 170 175

Ile Glu Asn Gly Glu Arg Arg Ser Tyr Thr Ile Lys Gln Gly His Val
180 185 190

Phe Val Ala Pro Ala Gly Ala Val Thr Tyr Leu Ala Asn Thr Asp Gly
195 200 205

Arg Lys Lys Leu Val Ile Thr Lys Ile Leu His Thr Ile Ser Val Pro
210 215 220

Gly Glu Phe Gln Phe Phe Phe Gly Pro Gly Gly Arg Asn Pro Glu Ser
225 230 235 240

Phe Leu Ser Ser Phe Ser Lys Ser Ile Gln Arg Ala Ala Tyr Lys Thr
245 250 255

Ser Ser Asp Arg Leu Glu Arg Leu Phe Gly Arg His Gly Gln Asp Lys
260 265 270

Gly Ile Ile Val Arg Ala Thr Glu Glu Gln Thr Arg Glu Leu Arg Arg
275 280 285

His Ala Ser Glu Gly Gly His Gly Pro His Trp Pro Leu Pro Pro Phe
290 295 300

Gly Glu Ser Arg Gly Pro Tyr Ser Leu Leu Asp Gln Arg Pro Ser Ile
305 310 315 320

Ala Asn Gln His Gly Gln Leu Tyr Glu Ala Asp Ala Arg Ser Phe His
325 330 335

Asp Leu Ala Glu His Asp Val Ser Val Ser Phe Ala Asn Ile Thr Ala
340 345 350

Gly Ser Met Ser Ala Pro Leu Tyr Asn Thr Arg Ser Phe Lys Ile Ala
355 360 365

Tyr Val Pro Asn Gly Lys Gly Tyr Ala Glu Ile Val Cys Pro His Arg
370 375 380

Gln Ser Gln Gly Gly Glu Ser Glu Arg Glu Arg Gly Lys Gly Arg Arg
385 390 395 400

Ser Glu Glu Glu Glu Glu Ser Ser Glu Glu Gln Glu Glu Val Gly Gln
405 410 415

Gly Tyr His Thr Ile Arg Ala Arg Leu Ser Pro Gly Thr Ala Phe Val
420 425 430

Val Pro Ala Gly His Pro Phe Val Ala Val Ala Ser Arg Asp Ser Asn
435 440 445

Leu Gln Ile Val Cys Phe Glu Val His Ala Asp Arg Asn Glu Lys Val
450 455 460

Phe Leu Ala Gly Ala Asp Asn Val Leu Gln Lys Leu Asp Arg Val Ala
465 470 475 480

Lys Ala Leu Ser Phe Ala Ser Lys Ala Glu Glu Val Asp Glu Val Leu
485 490 495

Gly Ser Arg Arg Glu Lys Gly Phe Leu Pro Gly Pro Lys Glu Ser Gly
500 505 510

Gly His Glu Glu Arg Glu Gln Glu Glu Glu Glu Arg Glu Glu Arg His
515 520 525

Gly Gly Arg Gly Glu Arg Glu Arg His Gly Arg Glu Glu Arg Glu Lys
530 535 540

Glu Glu Glu Glu Arg Glu Gly Arg His Gly Arg Gly Arg Arg Glu Glu
545 550 555 560

Val Ala Glu Thr Leu Leu Arg Met Val Thr Ala Arg Met
565 570

<210> 3
<211> 2003
<212> DNA
<213> Zea mays

<220>
<221> misc_feature
<223> Globulin-2 precursor

<400> 3
cgcacacacc cgagcatatc acagtgacac tacacgatgg tgagcgccag aatcgttgtc 60

ctcctcgccg tcctcctatg cgctgccgcc gcagtcgcgt cgtcctggga ggacgacaac 120

caccaccacc acgggggcca caagtccggg cgatgcgtgc ggcggtgcga ggaccggccc 180

tggcaccagc gcccccggtg cctggagcag tgcagggagg aggagcggga gaagcggcaa 240

gagcggagca ggcacgaggc cgacgaccgc agcggcgagg gctcgtcgga ggatgagcgc 300

gagcgcgagc aggagaagga ggagaagcag aaggaccggc ggccgtacgt gttcgaccgg 360

cgcagctttc gtcgcgtggt ccggagcgag caggggtccc tgagggtgct ccggccgttc 420

gacgaggtgt ccaggctcct ccgcggcatc cgggactacc gcgtggcggt cctggaggcg 480

aacccgcgct cgttcgtggt gcccagccac accgacgcgc actgcatcgg ctacgtggcg 540

gaaggcgagg gagtggtgac gacgatcgag aacggcgaga ggcggtcgta caccatcaag 600

caaggccacg tcttcgtggc gccggccggg gcggtcacct acctggccaa caccgacggc 660

cggaagaaac tggtcatcac caagatcctc cataccatct ccgtgcctgg cgagttccag 720

ttcttcttcg gccccggcgg gaggaacccg gaatcgttcc tgtcgagctt cagcaagagc 780

atccagagag ctgcgtacaa gacctcgagc gaccggctgg agaggctgtt cgggaggcat 840

gggcaggaca aggggatcat cgtgcgtgcc acggaggagc agacccgcga gctgcggcgc 900

cacgcctcgg agggcggcca cggcccgcac tggcccctgc cgccgttcgg cgagtcgcgc 960

ggcccctaca gcctcctgga ccagcggccc agcatcgcca accagcacgg gcagctctac 1020

gaggccgacg cgcgcagctt ccacgacctc gccgagcacg acgtcagcgt ctccttcgcc 1080

aacatcaccg cggggtccat gagcgcgcca ttgtacaaca cccgttcgtt caagatcgcc 1140

tacgtgccga acggcaaggg ctacgccgag atcgtgtgcc cgcaccgcca gtcgcagggc 1200

ggcgagagcg agcgcgagcg cggcaagggc aggaggagcg aagaagaaga agaatcgtct 1260

gaggagcagg aggaagtcgg gcaggggtac cacaccatcc gggcgcggct gtcaccgggc 1320

acggcgttcg tggtgcccgc gggccacccg ttcgtcgcgg tggcgtcccg ggacagcaac 1380

ctccagatcg tgtgcttcga ggtccacgcc gacaggaacg agaaggtgtt cctggccggc 1440

gccgacaacg tgctgcagaa gctcgaccgg gtcgccaagg cgctgtcatt cgcctccaag 1500

gcggaggagg tggacgaggt gctcggctcg cggcgcgaga aggggttcct tcctggcccc 1560

aaggagagcg gcggccacga ggagcgggag caggaggagg aggaacgcga agaacgccac 1620

ggcgggcgtg gggagaggga acgccacgga cgtgaggagc gggagaaaga ggaggaggaa 1680

cgcgaaggac gccacggccg cgggcgccgc gaggaagtgg cggagacgct cctgaggatg 1740

gtgaccgcca ggatgtgagg ccggccgtgc tcgccaaaac gagcaggaag caacgagagg 1800

gtggcgcgcg accgacgtgc gtacgtagca tgagcctgag tggagacgtt ggacgtgtat 1860

gtatatacct ctctgcgtgt taactatgta cgtaagcggc aggcagtgca ataagtgtgg 1920

ctctgtagta tgtacgtgcg ggtacgatgc tgtaagctac tgaggcaagt ccataaataa 1980

ataatgacac gtgcgtgttc tat 2003

<210> 4

<211> 450
<212> PRT
<213> Zea mays

<220>
<221> misc_feature
<223> Globulin-2 precursor

<400> 4

```
Met Lys Val Pro Val Leu Leu Leu Leu Val Ser Leu Cys Phe Ser Leu
1               5                   10                  15

Ala Leu Ala Trp Gln Thr Asp Thr Glu Ser Gly Ser Gly Arg Pro Tyr
            20                  25                  30

His Tyr Gly Glu Glu Ser Phe Arg His Trp Thr Arg Ser Arg Gln Gly
        35                  40                  45

Arg Phe Arg Val Leu Glu Arg Phe Thr His Glu Leu Leu Glu Asp Ala
    50                  55                  60

Val Gly Asn Tyr Arg Val Ala Glu Leu Glu Ala Ala Pro Arg Ala Phe
65                  70                  75                  80

Leu Gln Pro Ser His Tyr Asp Ala Asp Glu Val Met Phe Val Lys Glu
                85                  90                  95

Gly Glu Gly Val Ile Val Leu Leu Arg Gly Gly Lys Arg Glu Ser Phe
            100                 105                 110

Cys Val Arg Glu Gly Asp Val Met Val Ile Pro Ala Gly Ala Val Val
        115                 120                 125

Tyr Ser Ala Asn Thr His Gln Ser Glu Trp Phe Arg Val Val Met Leu
    130                 135                 140

Leu Ser Pro Val Val Ser Thr Ser Gly Arg Phe Glu Glu Phe Phe Pro
145                 150                 155                 160

Ile Gly Gly Glu Ser Pro Glu Ser Phe Leu Ser Val Phe Ser Asp Asp
                165                 170                 175

Val Ile Gln Ala Ser Phe Asn Thr Arg Arg Glu Glu Trp Glu Lys Val
            180                 185                 190

Phe Glu Lys Gln Ser Lys Gly Glu Ile Thr Thr Ala Ser Glu Glu Gln
        195                 200                 205
```

Ile Arg Glu Leu Ser Arg Ser Cys Ser Arg Gly Gly Arg Ser Ser Arg
210 215 220

Ser Glu Gly Gly Asp Ser Gly Ser Ser Ser Ser Lys Trp Glu Ile Lys
225 230 235 240

Pro Ser Ser Leu Thr Asp Lys Lys Pro Thr His Ser Asn Ser His Gly
245 250 255

Arg His Tyr Glu Ile Thr Gly Asp Glu Cys Pro His Leu Arg Leu Leu
260 265 270

Asp Met Asp Val Gly Leu Ala Asn Ile Ala Arg Gly Ser Met Met Ala
275 280 285

Pro Ser Tyr Asn Thr Arg Ala Asn Lys Ile Ala Ile Val Leu Lys Gly
290 295 300

Gln Gly Tyr Phe Glu Met Ala Cys Pro His Val Ser Gly Gly Arg Ser
305 310 315 320

Ser Pro Arg Arg Glu Arg Gly His Gly Arg Glu Glu Glu Glu Glu Arg
325 330 335

Glu Glu Glu Gln Gly Gly Gly Gly Gly Gln Lys Ser Arg Ser Tyr Arg
340 345 350

Gln Val Lys Ser Arg Ile Arg Glu Gly Ser Val Ile Val Ile Pro Ala
355 360 365

Gly His Pro Thr Ala Leu Val Ala Gly Glu Asp Lys Asn Leu Ala Val
370 375 380

Leu Cys Phe Glu Val Asn Ala Ser Phe Asp Asp Lys Val Phe Leu Ala
385 390 395 400

Gly Thr Asn Ser Ala Leu Gln Lys Met Asp Arg Pro Ala Lys Leu Leu
405 410 415

Ala Phe Gly Ala Asp Glu Glu Gln Gln Val Asp Arg Val Ile Gly Ala
420 425 430

Gln Lys Asp Ala Val Phe Leu Arg Gly Pro Gln Ser His Arg Val Ser
435 440 445

```
Ser Val
    450


<210>  5
<211>  723
<212>  DNA
<213>  Zea mays

<220>
<221>  CDS
<222>  (1)..(723)
<223>  Oleosin ZM-I


<400>  5
atg gca gcc aag att ttt gcc ctc ctt gcc ctc ctt gct ctt tca gca        48
Met Ala Ala Lys Ile Phe Ala Leu Leu Ala Leu Leu Ala Leu Ser Ala
1               5                   10                  15

aac gtt gct acc gcg act att att cca caa tgc tca caa caa tac ctc        96
Asn Val Ala Thr Ala Thr Ile Ile Pro Gln Cys Ser Gln Gln Tyr Leu
            20                  25                  30

tct ccg gtg aca gcc gcg aga ttt gaa tac cca act ata caa tcc tac       144
Ser Pro Val Thr Ala Ala Arg Phe Glu Tyr Pro Thr Ile Gln Ser Tyr
        35                  40                  45

agg cta caa cag gcc atc gca gca agc atc tta cgg tcg tta gca ttg       192
Arg Leu Gln Gln Ala Ile Ala Ala Ser Ile Leu Arg Ser Leu Ala Leu
    50                  55                  60

act gtc caa caa cca tat gcc cta ttg caa caa cca tcc tta gtg aat       240
Thr Val Gln Gln Pro Tyr Ala Leu Leu Gln Gln Pro Ser Leu Val Asn
65                  70                  75                  80

cta tat ctc caa aga atc gta gca caa caa cta caa caa caa ttg ctt       288
Leu Tyr Leu Gln Arg Ile Val Ala Gln Gln Leu Gln Gln Gln Leu Leu
                85                  90                  95

cca aca atc aat caa gta gtt gca gcg aac ctt gat gct tac ctc cag       336
Pro Thr Ile Asn Gln Val Val Ala Ala Asn Leu Asp Ala Tyr Leu Gln
            100                 105                 110

caa caa caa ttt ctt cca ttc aat caa cta gct ggg gtg aac cct gct       384
Gln Gln Gln Phe Leu Pro Phe Asn Gln Leu Ala Gly Val Asn Pro Ala
        115                 120                 125

gct tac ttg cag gca caa cag cta cta cca ttc aac caa ctt gtc agg       432
Ala Tyr Leu Gln Ala Gln Gln Leu Leu Pro Phe Asn Gln Leu Val Arg
    130                 135                 140

agc cct gct gcc ttc tta ctg cag caa cag ttg ttg cca ttc cat cta       480
Ser Pro Ala Ala Phe Leu Leu Gln Gln Gln Leu Leu Pro Phe His Leu
145                 150                 155                 160

caa gtt gtg gca aac att gct gct ttc ttg caa caa caa caa ttg ctg       528
Gln Val Val Ala Asn Ile Ala Ala Phe Leu Gln Gln Gln Gln Leu Leu
                165                 170                 175

cca ttt tac cca cag gtt gtg gga aac att aac gcc ttc ttg caa cag       576
```

Pro Phe Tyr Pro Gln Val Val Gly Asn Ile Asn Ala Phe Leu Gln Gln
180 185 190

caa cag ttg ctg cca ttc tac cca cag gat gtg gca aac aat gtc gcc 624
Gln Gln Leu Leu Pro Phe Tyr Pro Gln Asp Val Ala Asn Asn Val Ala
195 200 205

ttc tta caa caa caa caa ttg ctg cca ttt agc caa ctt gct ttg acg 672
Phe Leu Gln Gln Gln Gln Leu Leu Pro Phe Ser Gln Leu Ala Leu Thr
210 215 220

aat cct acc acc tta ttg cag cag ccc acc att ggt ggt gcc atc ttc 720
Asn Pro Thr Thr Leu Leu Gln Gln Pro Thr Ile Gly Gly Ala Ile Phe
225 230 235 240

tag 723

<210> 6
<211> 240
<212> PRT
<213> Zea mays

<400> 6

Met Ala Ala Lys Ile Phe Ala Leu Leu Ala Leu Leu Ala Leu Ser Ala
1 5 10 15

Asn Val Ala Thr Ala Thr Ile Ile Pro Gln Cys Ser Gln Gln Tyr Leu
20 25 30

Ser Pro Val Thr Ala Ala Arg Phe Glu Tyr Pro Thr Ile Gln Ser Tyr
35 40 45

Arg Leu Gln Gln Ala Ile Ala Ala Ser Ile Leu Arg Ser Leu Ala Leu
50 55 60

Thr Val Gln Gln Pro Tyr Ala Leu Leu Gln Gln Pro Ser Leu Val Asn
65 70 75 80

Leu Tyr Leu Gln Arg Ile Val Ala Gln Gln Leu Gln Gln Gln Leu Leu
85 90 95

Pro Thr Ile Asn Gln Val Val Ala Ala Asn Leu Asp Ala Tyr Leu Gln
100 105 110

Gln Gln Gln Phe Leu Pro Phe Asn Gln Leu Ala Gly Val Asn Pro Ala
115 120 125

Ala Tyr Leu Gln Ala Gln Gln Leu Leu Pro Phe Asn Gln Leu Val Arg
130 135 140

```
Ser Pro Ala Ala Phe Leu Leu Gln Gln Gln Leu Leu Pro Phe His Leu
145                 150                 155                 160

Gln Val Val Ala Asn Ile Ala Ala Phe Leu Gln Gln Gln Gln Leu Leu
                165                 170                 175

Pro Phe Tyr Pro Gln Val Val Gly Asn Ile Asn Ala Phe Leu Gln Gln
            180                 185                 190

Gln Gln Leu Leu Pro Phe Tyr Pro Gln Asp Val Ala Asn Asn Val Ala
        195                 200                 205

Phe Leu Gln Gln Gln Gln Leu Leu Pro Phe Ser Gln Leu Ala Leu Thr
    210                 215                 220

Asn Pro Thr Thr Leu Leu Gln Gln Pro Thr Ile Gly Gly Ala Ile Phe
225                 230                 235                 240



<210>  7
<211>  740
<212>  DNA
<213>  Zea mays

<220>
<221>  CDS
<222>  (79)..(534)
<223>  17.2 kD Heat Shock Protein


<400>  7
aacacgagcc cgaagcactc ttgcaatcca ctgagttctg tttgttgaga cgcatagagc      60

tagctgctag cgtcgaca atg tcg ctc gtg agg cgc agc aac gtg ttc gac       111
                    Met Ser Leu Val Arg Arg Ser Asn Val Phe Asp
                    1               5                   10

ccc ttc tcg atg gac ctc tgg gat ccc ttc gac acc atg ttc cgc tcc       159
Pro Phe Ser Met Asp Leu Trp Asp Pro Phe Asp Thr Met Phe Arg Ser
            15                  20                  25

atc gtc ccg tcg gcg acc tcc acc aac tcc gag act gcc gcc ttc gcc       207
Ile Val Pro Ser Ala Thr Ser Thr Asn Ser Glu Thr Ala Ala Phe Ala
        30                  35                  40

agc gcc cgc atc gac tgg aag gag acg ccc gag gcg cac gtc ttc aag       255
Ser Ala Arg Ile Asp Trp Lys Glu Thr Pro Glu Ala His Val Phe Lys
    45                  50                  55

gcc gac ctc ccc ggc gtc aag aag gag gag gtc aag gtt gag gtc gaa       303
Ala Asp Leu Pro Gly Val Lys Lys Glu Glu Val Lys Val Glu Val Glu
60                  65                  70                  75

gac ggc aac gtg ctg gtc atc agc ggc cag cgc agc agg gag aag gag       351
Asp Gly Asn Val Leu Val Ile Ser Gly Gln Arg Ser Arg Glu Lys Glu
```

80 85 90

gac aag gac gac aag tgg cac cgt gtc gag cgc agc agt ggc cag ttc 399
Asp Lys Asp Asp Lys Trp His Arg Val Glu Arg Ser Ser Gly Gln Phe
95 100 105

atc agg cgc ttc cgc ctg ccg gat gac gcc aag gtg gat cag gtc aag 447
Ile Arg Arg Phe Arg Leu Pro Asp Asp Ala Lys Val Asp Gln Val Lys
110 115 120

gct ggc ctc gag aac ggc gtg ctc acg gtc acc gtg cct aag gcg gaa 495
Ala Gly Leu Glu Asn Gly Val Leu Thr Val Thr Val Pro Lys Ala Glu
125 130 135

gag aag aag cct gag gtg aag gct att gag atc tct ggt tgagcatcca 544
Glu Lys Lys Pro Glu Val Lys Ala Ile Glu Ile Ser Gly
140 145 150

atccaatatg gacgtggatg aaggtgtact gctgctggtc cgtggctgtc gctgtcctgt 604

gtggatgttt cctgtatctt ctacagtata taatgtactt ccgtctgttt cgtttgtatg 664

tacaatctca atcttgcggg tatcgttcat gtatcccttt gaataataac aaataaaatc 724

gggtttgtca cggtaa 740

<210> 8
<211> 152
<212> PRT
<213> Zea mays

<400> 8

Met Ser Leu Val Arg Arg Ser Asn Val Phe Asp Pro Phe Ser Met Asp
1 5 10 15

Leu Trp Asp Pro Phe Asp Thr Met Phe Arg Ser Ile Val Pro Ser Ala
20 25 30

Thr Ser Thr Asn Ser Glu Thr Ala Ala Phe Ala Ser Ala Arg Ile Asp
35 40 45

Trp Lys Glu Thr Pro Glu Ala His Val Phe Lys Ala Asp Leu Pro Gly
50 55 60

Val Lys Lys Glu Glu Val Lys Val Glu Val Glu Asp Gly Asn Val Leu
65 70 75 80

Val Ile Ser Gly Gln Arg Ser Arg Glu Lys Glu Asp Lys Asp Asp Lys
85 90 95

Trp His Arg Val Glu Arg Ser Ser Gly Gln Phe Ile Arg Arg Phe Arg
100 105 110

```
Leu Pro Asp Asp Ala Lys Val Asp Gln Val Lys Ala Gly Leu Glu Asn
        115                 120                 125

Gly Val Leu Thr Val Thr Val Pro Lys Ala Glu Glu Lys Lys Pro Glu
    130                 135                 140

Val Lys Ala Ile Glu Ile Ser Gly
145                 150


<210>  9

<211>  469

<212>  DNA

<213>  Oryza sativa



<220>

<221>  CDS

<222>  (1)..(450)

<223>  17.2 kD Heat Shock Protein



<400>  9
atg tcg ctg gtg agg cgc agc aac gtg ttc gac cca ttc tcc ctc gac         48
Met Ser Leu Val Arg Arg Ser Asn Val Phe Asp Pro Phe Ser Leu Asp
1               5                   10                  15

ctc tgg gac ccc ttc gac agc gtg ttc cgc tcc gtc gtc ccg gcc acc         96
Leu Trp Asp Pro Phe Asp Ser Val Phe Arg Ser Val Val Pro Ala Thr
            20                  25                  30

tcc gac aac gac acc gcc gcc ttc gcc aac gcc cgc atc gac tgg aag        144
Ser Asp Asn Asp Thr Ala Ala Phe Ala Asn Ala Arg Ile Asp Trp Lys
        35                  40                  45

gag acg ccg gag tcg cac gtc ttc aag gcc gac ctc ccc ggc gtc aag        192
Glu Thr Pro Glu Ser His Val Phe Lys Ala Asp Leu Pro Gly Val Lys
    50                  55                  60

aag gag gag gtg aag gtg gag gtg gag gaa ggc aac gtg ctg gtg atc        240
Lys Glu Glu Val Lys Val Glu Val Glu Glu Gly Asn Val Leu Val Ile
65                  70                  75                  80

agc ggg cag cgc agc aag gag aag gag gac aag aac gac aag tgg cac        288
Ser Gly Gln Arg Ser Lys Glu Lys Glu Asp Lys Asn Asp Lys Trp His
                85                  90                  95

cgc gtg gag cgc agc agc ggg cag ttc atg cgg cgg ttc cgg ctg ccg        336
Arg Val Glu Arg Ser Ser Gly Gln Phe Met Arg Arg Phe Arg Leu Pro
            100                 105                 110
```

```
gag aac gcc aag gtg gac cag gtg aag gcc ggc atg gag aac ggc gtg      384
Glu Asn Ala Lys Val Asp Gln Val Lys Ala Gly Met Glu Asn Gly Val
        115                 120                 125

ctc acc gtc acc gtg ccc aag gcc gag gtc aag aag ccc gag gtg aag      432
Leu Thr Val Thr Val Pro Lys Ala Glu Val Lys Lys Pro Glu Val Lys
    130                 135                 140

gcc att gag atc tct ggc taaaatggtg aaaacggga                         469
Ala Ile Glu Ile Ser Gly
145                 150


<210>  10

<211>  150

<212>  PRT

<213>  Oryza sativa



<400>  10

Met Ser Leu Val Arg Arg Ser Asn Val Phe Asp Pro Phe Ser Leu Asp
1               5                   10                  15


Leu Trp Asp Pro Phe Asp Ser Val Phe Arg Ser Val Val Pro Ala Thr
            20                  25                  30


Ser Asp Asn Asp Thr Ala Ala Phe Ala Asn Ala Arg Ile Asp Trp Lys
        35                  40                  45


Glu Thr Pro Glu Ser His Val Phe Lys Ala Asp Leu Pro Gly Val Lys
    50                  55                  60


Lys Glu Glu Val Lys Val Glu Val Glu Glu Gly Asn Val Leu Val Ile
65                  70                  75                  80


Ser Gly Gln Arg Ser Lys Glu Lys Glu Asp Lys Asn Asp Lys Trp His
                85                  90                  95


Arg Val Glu Arg Ser Ser Gly Gln Phe Met Arg Arg Phe Arg Leu Pro
            100                 105                 110


Glu Asn Ala Lys Val Asp Gln Val Lys Ala Gly Met Glu Asn Gly Val
        115                 120                 125


Leu Thr Val Thr Val Pro Lys Ala Glu Val Lys Lys Pro Glu Val Lys
    130                 135                 140
```

Ala Ile Glu Ile Ser Gly
145 150

<210> 11

<211> 460

<212> DNA

<213> wheat

<220>

<221> CDS

<222> (1)..(453)

<223>

<400> 11
atg tcg atc gtg agg cgg acg aac gtg ttc gac ccc ttc gcc gac ctc 48
Met Ser Ile Val Arg Arg Thr Asn Val Phe Asp Pro Phe Ala Asp Leu
1 5 10 15

tgg gcg gac ccc ttc gac acc ttc cgc tcc atc gtc ccg gcg atc tca 96
Trp Ala Asp Pro Phe Asp Thr Phe Arg Ser Ile Val Pro Ala Ile Ser
20 25 30

ggc ggc ggc agc gag acg gct gcg ttc gcc aac gcc cgg atg gac tgg 144
Gly Gly Gly Ser Glu Thr Ala Ala Phe Ala Asn Ala Arg Met Asp Trp
35 40 45

aag gag acc ccc gaa gcg cac gtc ttc aag gcc gac ctc ccc ggc gtg 192
Lys Glu Thr Pro Glu Ala His Val Phe Lys Ala Asp Leu Pro Gly Val
50 55 60

aag aag gag gag gtc aag gtg gag gtg gag gac ggc aac gtg ctc gtc 240
Lys Lys Glu Glu Val Lys Val Glu Val Glu Asp Gly Asn Val Leu Val
65 70 75 80

gtc agc ggc gag cgt aca aag gag aag gag gac aag aac gac aag tgg 288
Val Ser Gly Glu Arg Thr Lys Glu Lys Glu Asp Lys Asn Asp Lys Trp
85 90 95

cac cgc gtg gag cgc agc agc ggc aag ttc gtg cgg cgc ttc cgg ctg 336
His Arg Val Glu Arg Ser Ser Gly Lys Phe Val Arg Arg Phe Arg Leu
100 105 110

ctg gag gac gcc aag gtg gag gag gtg aag gcc ggg ctg gag aac ggg 384
Leu Glu Asp Ala Lys Val Glu Glu Val Lys Ala Gly Leu Glu Asn Gly
115 120 125

gtg ctc acc gtc acc gtg ccc aag gcc gag gtc aag aag ccc gag gtg 432

Val Leu Thr Val Thr Val Pro Lys Ala Glu Val Lys Lys Pro Glu Val
130 135 140

aag gcc atc cag atc tcc ggc tgagtat 460
Lys Ala Ile Gln Ile Ser Gly
145 150

<210> 12

<211> 151

<212> PRT

<213> wheat

<400> 12

Met Ser Ile Val Arg Arg Thr Asn Val Phe Asp Pro Phe Ala Asp Leu
1 5 10 15

Trp Ala Asp Pro Phe Asp Thr Phe Arg Ser Ile Val Pro Ala Ile Ser
20 25 30

Gly Gly Gly Ser Glu Thr Ala Ala Phe Ala Asn Ala Arg Met Asp Trp
35 40 45

Lys Glu Thr Pro Glu Ala His Val Phe Lys Ala Asp Leu Pro Gly Val
50 55 60

Lys Lys Glu Glu Val Lys Val Glu Val Glu Asp Gly Asn Val Leu Val
65 70 75 80

Val Ser Gly Glu Arg Thr Lys Glu Lys Glu Asp Lys Asn Asp Lys Trp
85 90 95

His Arg Val Glu Arg Ser Ser Gly Lys Phe Val Arg Arg Phe Arg Leu
100 105 110

Leu Glu Asp Ala Lys Val Glu Glu Val Lys Ala Gly Leu Glu Asn Gly
115 120 125

Val Leu Thr Val Thr Val Pro Lys Ala Glu Val Lys Lys Pro Glu Val
130 135 140

Lys Ala Ile Gln Ile Ser Gly
145 150

<210> 13

```
<211>  876
<212>  DNA
<213>  Hordeum vulgare

<220>
<221>  CDS
<222>  (1)..(876)
<223>  PIR2 Accesion T06212
       Glucose and Ribitol Dehydrogenoase Homolog

<400>  13
atg gcg tcg cag aag ttc ccg ccg cag cag cag gac tgc cag ccc ggc       48
Met Ala Ser Gln Lys Phe Pro Pro Gln Gln Gln Asp Cys Gln Pro Gly
1               5                   10                  15

aag gag cac gcc atg gac ccc cgc ccc gag gcc atc atc aag aac tac       96
Lys Glu His Ala Met Asp Pro Arg Pro Glu Ala Ile Ile Lys Asn Tyr
            20                  25                  30

aag tcg ggc caa caa gct cca ggg caa ggt ggc gct ggt gac cgg cgg      144
Lys Ser Gly Gln Gln Ala Pro Gly Gln Gly Gly Ala Gly Asp Arg Arg
        35                  40                  45

cga ctc ggg cat cgg gcg cgc ggt gtg cct gtg cct cgc gct gga ggg      192
Arg Leu Gly His Arg Ala Arg Gly Val Pro Val Pro Arg Ala Gly Gly
    50                  55                  60

cgc gac ggt gaa ctt cac gta cgt gaa ggg gca cga gga caa gga cgc      240
Arg Asp Gly Glu Leu His Val Arg Glu Gly Ala Arg Gly Gln Gly Arg
65                  70                  75                  80

gga gga gac cct gca ggc gct ccg cga cat caa gtc ccg cac cgg cgc      288
Gly Gly Asp Pro Ala Gly Ala Pro Arg His Gln Val Pro His Arg Arg
                85                  90                  95

cgg cga gcc caa ggc gct ctc ggg cga cct cgg gta cga gga gaa ctg      336
Arg Arg Ala Gln Gly Ala Leu Gly Arg Pro Arg Val Arg Gly Glu Leu
            100                 105                 110

ccg cag ggt ggt gga gga ggt ggc caa cgc gca cgg cgg ccg cgt gga      384
Pro Gln Gly Gly Gly Gly Gly Gly Gln Arg Ala Arg Arg Pro Arg Gly
        115                 120                 125

cat cct cgt gaa caa cgc ggc cga gca gta cgt ccg ccc ctg cat cac      432
His Pro Arg Glu Gln Arg Gly Arg Ala Val Arg Pro Pro Leu His His
    130                 135                 140

cga gat cac cga gca gga cct gga gcg cgt gtt ccg cac caa cat ctt      480
Arg Asp His Arg Ala Gly Pro Gly Ala Arg Val Pro His Gln His Leu
145                 150                 155                 160

ctc cta ctt cct cat gac caa gtt cgc cgt gaa gca cat ggg gcc cgg      528
Leu Leu Leu Pro His Asp Gln Val Arg Arg Glu Ala His Gly Ala Arg
                165                 170                 175

gtc cag cat cat caa cac cac ctc cgt gaa cgc gta caa ggg caa cgc      576
Val Gln His His Gln His His Leu Arg Glu Arg Val Gln Gly Gln Arg
            180                 185                 190

gac gct gct gga cta cac ggc cac caa ggg cgc cat cgt ggc ctt cac      624
Asp Ala Ala Gly Leu His Gly His Gln Gly Arg His Arg Gly Leu His
```

195 200 205

ccg cgc gct gtc gat gca gct ggc gga gaa ggg gat ccg cgt caa cgg 672
Pro Arg Ala Val Asp Ala Ala Gly Gly Glu Gly Asp Pro Arg Gln Arg
210 215 220

cgt ggc gcc ggg gcc cat ctg gac gcc cct cat ccc ggc ctc ctt ccc 720
Arg Gly Ala Gly Ala His Leu Asp Ala Pro His Pro Gly Leu Leu Pro
225 230 235 240

gga gga gaa ggt gaa gca gtt cgg gtc cga ggt gcc cat gaa gcg cgc 768
Gly Gly Glu Gly Glu Ala Val Arg Val Arg Gly Ala His Glu Ala Arg
245 250 255

cat gca gcc cag cga ggt cgc gcc cag ctt cgt ctt cct tgc cag cga 816
His Ala Ala Gln Arg Gly Arg Ala Gln Leu Arg Leu Pro Cys Gln Arg
260 265 270

gca gga ctc ctc cta cat ctc cgg cca gat cct cca ccc caa cgg tgg 864
Ala Gly Leu Leu Leu His Leu Arg Pro Asp Pro Pro Pro Gln Arg Trp
275 280 285

tac cat cgt caa 876
Tyr His Arg Gln
290

<210> 14
<211> 292
<212> PRT
<213> Hordeum vulgare

<400> 14

Met Ala Ser Gln Lys Phe Pro Pro Gln Gln Gln Asp Cys Gln Pro Gly
1 5 10 15

Lys Glu His Ala Met Asp Pro Arg Pro Glu Ala Ile Ile Lys Asn Tyr
20 25 30

Lys Ser Gly Gln Gln Ala Pro Gly Gln Gly Gly Ala Gly Asp Arg Arg
35 40 45

Arg Leu Gly His Arg Ala Arg Gly Val Pro Val Pro Arg Ala Gly Gly
50 55 60

Arg Asp Gly Glu Leu His Val Arg Glu Gly Ala Arg Gly Gln Gly Arg
65 70 75 80

Gly Gly Asp Pro Ala Gly Ala Pro Arg His Gln Val Pro His Arg Arg
85 90 95

Arg Arg Ala Gln Gly Ala Leu Gly Arg Pro Arg Val Arg Gly Glu Leu
100 105 110

Pro Gln Gly Gly Gly Gly Gly Gly Gln Arg Ala Arg Arg Pro Arg Gly
115 120 125

His Pro Arg Glu Gln Arg Gly Arg Ala Val Arg Pro Pro Leu His His
130 135 140

Arg Asp His Arg Ala Gly Pro Gly Ala Arg Val Pro His Gln His Leu
145 150 155 160

Leu Leu Leu Pro His Asp Gln Val Arg Arg Glu Ala His Gly Ala Arg
165 170 175

Val Gln His His Gln His His Leu Arg Glu Arg Val Gln Gly Gln Arg
180 185 190

Asp Ala Ala Gly Leu His Gly His Gln Gly Arg His Arg Gly Leu His
195 200 205

Pro Arg Ala Val Asp Ala Ala Gly Gly Glu Gly Asp Pro Arg Gln Arg
210 215 220

Arg Gly Ala Gly Ala His Leu Asp Ala Pro His Pro Gly Leu Leu Pro
225 230 235 240

Gly Gly Glu Gly Glu Ala Val Arg Val Arg Gly Ala His Glu Ala Arg
245 250 255

His Ala Ala Gln Arg Gly Arg Ala Gln Leu Arg Leu Pro Cys Gln Arg
260 265 270

Ala Gly Leu Leu Leu His Leu Arg Pro Asp Pro Pro Pro Gln Arg Trp
275 280 285

Tyr His Arg Gln
290

<210> 15
<211> 1187

<212> DNA

<213> Soybean

<220>

<221> CDS

<222> (38)..(916)

<223> Glucose and Ribitol Dehydrogenoase Homolog

<400> 15

```
gatctccata tctgttccca aaagcttgtt tgtaaga atg gct tcc ggt gaa cag      55
                                         Met Ala Ser Gly Glu Gln
                                         1               5

aaa ttc cct cct caa caa caa caa aca cag cct ggg aag gag cat gct      103
Lys Phe Pro Pro Gln Gln Gln Gln Thr Gln Pro Gly Lys Glu His Ala
            10                  15                  20

atg act cca gta ccc caa ttc act agc cct gac tac aag cct tca aat      151
Met Thr Pro Val Pro Gln Phe Thr Ser Pro Asp Tyr Lys Pro Ser Asn
        25                  30                  35

aaa ctt caa ggg aag att gca tta gtc act ggg ggt gat tct ggg att      199
Lys Leu Gln Gly Lys Ile Ala Leu Val Thr Gly Gly Asp Ser Gly Ile
    40                  45                  50

gga cga gcg gtg tgt aac ttg ttt gcc tta gaa ggt gct acc gtg gcc      247
Gly Arg Ala Val Cys Asn Leu Phe Ala Leu Glu Gly Ala Thr Val Ala
55                  60                  65                  70

ttc acg tat gtg aag ggg cat gag gac aag gac gcg agg gac aca ttg      295
Phe Thr Tyr Val Lys Gly His Glu Asp Lys Asp Ala Arg Asp Thr Leu
                75                  80                  85

gaa atg atc aag aga gca aag act tcg gat gcc aag gat cca atg gca      343
Glu Met Ile Lys Arg Ala Lys Thr Ser Asp Ala Lys Asp Pro Met Ala
            90                  95                  100

ata gca tct gat ttg ggt tac gat gag aac tgc aag agg gtg gtt gat      391
Ile Ala Ser Asp Leu Gly Tyr Asp Glu Asn Cys Lys Arg Val Val Asp
        105                 110                 115

gag gtc gtg agt gct tat ggt tgt att gac att ctg gtc aac aat gca      439
Glu Val Val Ser Ala Tyr Gly Cys Ile Asp Ile Leu Val Asn Asn Ala
    120                 125                 130

gct gag cag tac gag tgt gga acc gtg gag gac ata gac gag cct agg      487
Ala Glu Gln Tyr Glu Cys Gly Thr Val Glu Asp Ile Asp Glu Pro Arg
135                 140                 145                 150

ctt gag agg gtc ttt cgt aca aat atc ttc tcc tat ttc ttc atg gcg      535
Leu Glu Arg Val Phe Arg Thr Asn Ile Phe Ser Tyr Phe Phe Met Ala
                155                 160                 165

agg cat gcc ttg aag cac atg aag gaa gga agc agc att atc aac acg      583
Arg His Ala Leu Lys His Met Lys Glu Gly Ser Ser Ile Ile Asn Thr
            170                 175                 180
```

```
aca tca gtg aat gca tac aag gga cat gcg aaa cta ttg gac tac acg      631
Thr Ser Val Asn Ala Tyr Lys Gly His Ala Lys Leu Leu Asp Tyr Thr
        185                 190                 195

tcc acc aag ggg gca att gtg gcc tat aca agg ggt ctt gcc ctt cag      679
Ser Thr Lys Gly Ala Ile Val Ala Tyr Thr Arg Gly Leu Ala Leu Gln
    200                 205                 210

ctg gtg agt aag gga att cgg gtt aat ggg gtg gct cca ggg ccc att      727
Leu Val Ser Lys Gly Ile Arg Val Asn Gly Val Ala Pro Gly Pro Ile
215                 220                 225                 230

tgg acc cct ttg ata cca gcc tct ttc aag gag gaa gaa acg gcc caa      775
Trp Thr Pro Leu Ile Pro Ala Ser Phe Lys Glu Glu Glu Thr Ala Gln
                235                 240                 245

ttt gga gcg cag gtc cca atg aag aga gct ggt caa cct att gag gtt      823
Phe Gly Ala Gln Val Pro Met Lys Arg Ala Gly Gln Pro Ile Glu Val
            250                 255                 260

gct cct tcc tat gtt ttt ctt gct tcc aac caa tgc tcc tct tac ata      871
Ala Pro Ser Tyr Val Phe Leu Ala Ser Asn Gln Cys Ser Ser Tyr Ile
        265                 270                 275


act gga caa gtc ctt cac ccc aat ggt gga acc gtt gtg aat ggt          916
Thr Gly Gln Val Leu His Pro Asn Gly Gly Thr Val Val Asn Gly
    280                 285                 290

taaaccgttg gtgatgatga tattcgggat gaatatatgt ggcgagagta gtaggccagt    976

gttacgtttt gtgtgaatgt tttacgatgt gttttaatgc atggctaact cactcaggtc   1036

ctctctgcac tgttagaggt ggggcttgga ggattatcca cttttgaatg tacgagttat   1096

tagcctaaga aaatgtgtct tttgtagcca attatatgta aacaagtaaa agtatataat   1156

aaagatcggt atgtataagg tttaaacttt a                                  1187
```

<210> 16

<211> 293

<212> PRT

<213> Soybean

<400> 16

```
Met Ala Ser Gly Glu Gln Lys Phe Pro Pro Gln Gln Gln Gln Thr Gln
1               5                   10                  15

Pro Gly Lys Glu His Ala Met Thr Pro Val Pro Gln Phe Thr Ser Pro
            20                  25                  30

Asp Tyr Lys Pro Ser Asn Lys Leu Gln Gly Lys Ile Ala Leu Val Thr
```

35 40 45

Gly Gly Asp Ser Gly Ile Gly Arg Ala Val Cys Asn Leu Phe Ala Leu
50 55 60

Glu Gly Ala Thr Val Ala Phe Thr Tyr Val Lys Gly His Glu Asp Lys
65 70 75 80

Asp Ala Arg Asp Thr Leu Glu Met Ile Lys Arg Ala Lys Thr Ser Asp
85 90 95

Ala Lys Asp Pro Met Ala Ile Ala Ser Asp Leu Gly Tyr Asp Glu Asn
100 105 110

Cys Lys Arg Val Val Asp Glu Val Val Ser Ala Tyr Gly Cys Ile Asp
115 120 125

Ile Leu Val Asn Asn Ala Ala Glu Gln Tyr Glu Cys Gly Thr Val Glu
130 135 140

Asp Ile Asp Glu Pro Arg Leu Glu Arg Val Phe Arg Thr Asn Ile Phe
145 150 155 160

Ser Tyr Phe Phe Met Ala Arg His Ala Leu Lys His Met Lys Glu Gly
165 170 175

Ser Ser Ile Ile Asn Thr Thr Ser Val Asn Ala Tyr Lys Gly His Ala
180 185 190

Lys Leu Leu Asp Tyr Thr Ser Thr Lys Gly Ala Ile Val Ala Tyr Thr
195 200 205

Arg Gly Leu Ala Leu Gln Leu Val Ser Lys Gly Ile Arg Val Asn Gly
210 215 220

Val Ala Pro Gly Pro Ile Trp Thr Pro Leu Ile Pro Ala Ser Phe Lys
225 230 235 240

Glu Glu Glu Thr Ala Gln Phe Gly Ala Gln Val Pro Met Lys Arg Ala
245 250 255

Gly Gln Pro Ile Glu Val Ala Pro Ser Tyr Val Phe Leu Ala Ser Asn
260 265 270

Gln Cys Ser Ser Tyr Ile Thr Gly Gln Val Leu His Pro Asn Gly Gly
275 280 285

Thr Val Val Asn Gly
290

<210> 17
<211> 1194

<212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS

<222> (119)..(982)

<223> PIR2 Accesion T06212 Glucose and Ribitol Dehydrogenase Homolog

<400> 17
cacttcgtca agagtcttat catcaaacgt tacgttctcg ttttctcaaa ttccaaacaa 60

taccaagaaa cctctacttg aaaagagaag attatcgccg cgtgtgtgcc taagagcg 118

atg gcg tct gag aaa caa aaa caa cat gca caa cct ggc aaa gaa cat 166
Met Ala Ser Glu Lys Gln Lys Gln His Ala Gln Pro Gly Lys Glu His
1 5 10 15

gtc atg gaa tca agc cca caa ttc tct agc tca gat tac caa cct tcc 214
Val Met Glu Ser Ser Pro Gln Phe Ser Ser Ser Asp Tyr Gln Pro Ser
20 25 30

aac aag ctt cgt ggt aag gtg gcg ttg ata act ggt gga gac tct ggg 262
Asn Lys Leu Arg Gly Lys Val Ala Leu Ile Thr Gly Gly Asp Ser Gly
35 40 45

att ggt cga gcc gtg gga tac tgt ttt gca tcc gaa gga gct act gtg 310
Ile Gly Arg Ala Val Gly Tyr Cys Phe Ala Ser Glu Gly Ala Thr Val
50 55 60

gct ttc act tac gtg aag ggt caa gaa gaa aaa gat gca caa gag acc 358
Ala Phe Thr Tyr Val Lys Gly Gln Glu Glu Lys Asp Ala Gln Glu Thr
65 70 75 80

cta caa atg ttg aag gag gtc aaa acc tcg gac tcc aag gaa cct atc 406
Leu Gln Met Leu Lys Glu Val Lys Thr Ser Asp Ser Lys Glu Pro Ile
85 90 95

gcc att cca acg gat tta gga ttt gac gaa aac tgc aaa agg gtc gtt 454
Ala Ile Pro Thr Asp Leu Gly Phe Asp Glu Asn Cys Lys Arg Val Val
100 105 110

gat gag gtc gtt aat gct ttt ggc cgc atc gat gtt ttg atc aat aac 502

```
Asp Glu Val Val Asn Ala Phe Gly Arg Ile Asp Val Leu Ile Asn Asn
        115                 120                 125

gca gca gag cag tac gag agc agc aca atc gaa gag att gat gag cct      550
Ala Ala Glu Gln Tyr Glu Ser Ser Thr Ile Glu Glu Ile Asp Glu Pro
    130                 135                 140

agg ctt gag cga gtc ttc cgt aca aac atc ttt tct tac ttc ttt ctc      598
Arg Leu Glu Arg Val Phe Arg Thr Asn Ile Phe Ser Tyr Phe Phe Leu
145                 150                 155                 160

aca agg cat gcg ttg aag cat atg aag gaa gga agc agc att atc aac      646
Thr Arg His Ala Leu Lys His Met Lys Glu Gly Ser Ser Ile Ile Asn
                165                 170                 175

acc act tcg gtg aat gcc tac aag gga aac gct tca ctt ctc gac tac      694
Thr Thr Ser Val Asn Ala Tyr Lys Gly Asn Ala Ser Leu Leu Asp Tyr
            180                 185                 190

acc gct aca aaa gga gcg att gtg gcg ttt act cga gga ctt gca ctt      742
Thr Ala Thr Lys Gly Ala Ile Val Ala Phe Thr Arg Gly Leu Ala Leu
        195                 200                 205

cag cta gct gag aaa gga atc cgt gtc aat ggt gtg gct cct ggt cca      790
Gln Leu Ala Glu Lys Gly Ile Arg Val Asn Gly Val Ala Pro Gly Pro
    210                 215                 220

ata tgg aca ccc ctt atc cca gca tca ttc aat gag gag aag att aag      838
Ile Trp Thr Pro Leu Ile Pro Ala Ser Phe Asn Glu Glu Lys Ile Lys
225                 230                 235                 240

aat ttt ggg tct gag gtt ccg atg aaa aga gcg ggt cag cca att gaa      886
Asn Phe Gly Ser Glu Val Pro Met Lys Arg Ala Gly Gln Pro Ile Glu
                245                 250                 255

gtg gca cca tcc tat gtt ttc ttg gcg tgt aac cac tgc tct tct tac      934
Val Ala Pro Ser Tyr Val Phe Leu Ala Cys Asn His Cys Ser Ser Tyr
            260                 265                 270

ttc act ggt caa gtt ctt cac cct aat gga gga gct gtg gta aat gcg      982
Phe Thr Gly Gln Val Leu His Pro Asn Gly Gly Ala Val Val Asn Ala
        275                 280                 285

taagcgtgga gtggacaaga ccggtcttaa cgtcttagac cattaaatat gatatgatga   1042

tgttgtttga gtttaggggc tttttgttat gttggtaatg tgttacgtcc gtatatgttg   1102

gtaatgtgtt gcgtccgtac cttctgtagc aaaagtatgt gtttaataaa gactttacct   1162

cgacaatagt cggggcctgt tcccttaaaa gt                                  1194
```

<210> 18

<211> 288

<212> PRT

<213> Arabidopsis thaliana

<400> 18

```
Met Ala Ser Glu Lys Gln Lys Gln His Ala Gln Pro Gly Lys Glu His
1               5                   10                  15

Val Met Glu Ser Ser Pro Gln Phe Ser Ser Ser Asp Tyr Gln Pro Ser
            20                  25                  30

Asn Lys Leu Arg Gly Lys Val Ala Leu Ile Thr Gly Gly Asp Ser Gly
        35                  40                  45

Ile Gly Arg Ala Val Gly Tyr Cys Phe Ala Ser Glu Gly Ala Thr Val
    50                  55                  60

Ala Phe Thr Tyr Val Lys Gly Gln Glu Glu Lys Asp Ala Gln Glu Thr
65                  70                  75                  80

Leu Gln Met Leu Lys Glu Val Lys Thr Ser Asp Ser Lys Glu Pro Ile
                85                  90                  95

Ala Ile Pro Thr Asp Leu Gly Phe Asp Glu Asn Cys Lys Arg Val Val
            100                 105                 110

Asp Glu Val Val Asn Ala Phe Gly Arg Ile Asp Val Leu Ile Asn Asn
        115                 120                 125

Ala Ala Glu Gln Tyr Glu Ser Ser Thr Ile Glu Glu Ile Asp Glu Pro
    130                 135                 140

Arg Leu Glu Arg Val Phe Arg Thr Asn Ile Phe Ser Tyr Phe Phe Leu
145                 150                 155                 160

Thr Arg His Ala Leu Lys His Met Lys Glu Gly Ser Ser Ile Ile Asn
                165                 170                 175

Thr Thr Ser Val Asn Ala Tyr Lys Gly Asn Ala Ser Leu Leu Asp Tyr
            180                 185                 190

Thr Ala Thr Lys Gly Ala Ile Val Ala Phe Thr Arg Gly Leu Ala Leu
        195                 200                 205

Gln Leu Ala Glu Lys Gly Ile Arg Val Asn Gly Val Ala Pro Gly Pro
    210                 215                 220

Ile Trp Thr Pro Leu Ile Pro Ala Ser Phe Asn Glu Glu Lys Ile Lys
225                 230                 235                 240
```

```
Asn Phe Gly Ser Glu Val Pro Met Lys Arg Ala Gly Gln Pro Ile Glu
                245                 250                 255

Val Ala Pro Ser Tyr Val Phe Leu Ala Cys Asn His Cys Ser Ser Tyr
            260                 265                 270

Phe Thr Gly Gln Val Leu His Pro Asn Gly Gly Ala Val Val Asn Ala


<210>  19
<211>  2451
<212>  DNA
<213>  Zea mays

<220>
<221>  CDS
<222>  (1)..(2451)
<223>  RECEPTOR PROTEIN KINASE ZMPK1 PRECURSOR (EC 2.7.1.-) -


<400>  19
atg cct cgt cct ctt gca gct ctc ctc tct acc gct tgc atc ctc tcc        48
Met Pro Arg Pro Leu Ala Ala Leu Leu Ser Thr Ala Cys Ile Leu Ser
1               5                   10                  15

ttc ttc atc gct ctg ttt cca aga gct gca tca tcc cga gac atc cta        96
Phe Phe Ile Ala Leu Phe Pro Arg Ala Ala Ser Ser Arg Asp Ile Leu
            20                  25                  30

cca ctg ggt tcc tct ctc gta gtc gag tcc tac gaa tcc agc acc tta       144
Pro Leu Gly Ser Ser Leu Val Val Glu Ser Tyr Glu Ser Ser Thr Leu
        35                  40                  45

caa tca tca gac ggg aca ttc tcc tct ggc ttc tac gaa gtc tac acc       192
Gln Ser Ser Asp Gly Thr Phe Ser Ser Gly Phe Tyr Glu Val Tyr Thr
    50                  55                  60

cat gcc ttc aca ttc tca gta tgg tac tca aag acg gag gcg gcg gcg       240
His Ala Phe Thr Phe Ser Val Trp Tyr Ser Lys Thr Glu Ala Ala Ala
65                  70                  75                  80

gcc aac aac aag acc atc gtg tgg agc gca aac cct gac cgc cct gtc       288
Ala Asn Asn Lys Thr Ile Val Trp Ser Ala Asn Pro Asp Arg Pro Val
                85                  90                  95

cat gcc agg agg tcg gct cta acc ctg caa aag gac ggc aac atg gtg       336
His Ala Arg Arg Ser Ala Leu Thr Leu Gln Lys Asp Gly Asn Met Val
            100                 105                 110

ctc acc gac tac gac ggc gca gcc gtg tgg cga gct gat ggc aac aac       384
Leu Thr Asp Tyr Asp Gly Ala Ala Val Trp Arg Ala Asp Gly Asn Asn
        115                 120                 125

ttc acc ggc gtc cag cgt gct cgg ctc ctg gac acc ggg aac ctc gtc       432
Phe Thr Gly Val Gln Arg Ala Arg Leu Leu Asp Thr Gly Asn Leu Val
    130                 135                 140
```

atc gag gac tca gga ggt aac act gta tgg cag agt ttc gat tcc cca 480
Ile Glu Asp Ser Gly Gly Asn Thr Val Trp Gln Ser Phe Asp Ser Pro
145 150 155 160

acg gac act ttc ctg ccg acg cag ctc atc act gct gcg acc aga tta 528
Thr Asp Thr Phe Leu Pro Thr Gln Leu Ile Thr Ala Ala Thr Arg Leu
165 170 175

gtc ccc aca acc caa tcg cgt agt cct ggt aac tac atc ttc cgc ttc 576
Val Pro Thr Thr Gln Ser Arg Ser Pro Gly Asn Tyr Ile Phe Arg Phe
180 185 190

agc gac ctc tca gtg ctg tcg ctt ata tac cac gtg cct caa gtc tca 624
Ser Asp Leu Ser Val Leu Ser Leu Ile Tyr His Val Pro Gln Val Ser
195 200 205

gac ata tac tgg cca gac cct gac cag aac ctc tac cag gat ggc cgg 672
Asp Ile Tyr Trp Pro Asp Pro Asp Gln Asn Leu Tyr Gln Asp Gly Arg
210 215 220

aac cag tat aac agt acg agg tta gga atg ctt act gat agc ggg gtg 720
Asn Gln Tyr Asn Ser Thr Arg Leu Gly Met Leu Thr Asp Ser Gly Val
225 230 235 240

ctt gcc tcg agc gac ttc gct gat ggt cag gcg ctt gtg gcc tcc gac 768
Leu Ala Ser Ser Asp Phe Ala Asp Gly Gln Ala Leu Val Ala Ser Asp
245 250 255

gta ggg ccg ggc gtc aag aga agg cta act ctt gac cct gat ggc aat 816
Val Gly Pro Gly Val Lys Arg Arg Leu Thr Leu Asp Pro Asp Gly Asn
260 265 270

ctc cgt ctg tac agc atg aac gat tca gat ggg tca tgg tca gtt tca 864
Leu Arg Leu Tyr Ser Met Asn Asp Ser Asp Gly Ser Trp Ser Val Ser
275 280 285

atg gta gca atg acc cag cct tgc aat att cac ggt ttg tgt ggt cct 912
Met Val Ala Met Thr Gln Pro Cys Asn Ile His Gly Leu Cys Gly Pro
290 295 300

aat ggc atc tgc cac tac tca ccc aca cct aca tgt tcg tgc cca cca 960
Asn Gly Ile Cys His Tyr Ser Pro Thr Pro Thr Cys Ser Cys Pro Pro
305 310 315 320

ggt tat gcg acg agg aac ccg ggt aac tgg act gaa ggc tgt atg gct 1008
Gly Tyr Ala Thr Arg Asn Pro Gly Asn Trp Thr Glu Gly Cys Met Ala
325 330 335

att gtc aac aca acc tgt gac cgc tat gac aag agg tct atg aga ttt 1056
Ile Val Asn Thr Thr Cys Asp Arg Tyr Asp Lys Arg Ser Met Arg Phe
340 345 350

gtg cga ctt ccc aat acg gat ttt tgg ggg tcg gat cag caa cat ctt 1104
Val Arg Leu Pro Asn Thr Asp Phe Trp Gly Ser Asp Gln Gln His Leu
355 360 365

ctg tcg gtt tct ctt cga act tgt agg gat atc tgc atc agt gac tgc 1152
Leu Ser Val Ser Leu Arg Thr Cys Arg Asp Ile Cys Ile Ser Asp Cys
370 375 380

acc tgt aaa ggc ttt cag tat cag gaa ggc aca gga tca tgc tat cca 1200

Thr Cys Lys Gly Phe Gln Tyr Gln Glu Gly Thr Gly Ser Cys Tyr Pro
385 390 395 400

aaa gct tat ctt ttc agt gga aga acc tac cca aca tct gac gtg cga 1248
Lys Ala Tyr Leu Phe Ser Gly Arg Thr Tyr Pro Thr Ser Asp Val Arg
405 410 415

acg ata tat ctc aag ctt cca aca ggg gtc agt gtt tca aat gcc ctt 1296
Thr Ile Tyr Leu Lys Leu Pro Thr Gly Val Ser Val Ser Asn Ala Leu
420 425 430

att cca cgt tcc gac gtg ttc gat tcc gtg ccc cgt cgt ctc gac tgc 1344
Ile Pro Arg Ser Asp Val Phe Asp Ser Val Pro Arg Arg Leu Asp Cys
435 440 445

gat cgg atg aac aaa agc atc aga gaa ccg ttt cca gat gtg cac aag 1392
Asp Arg Met Asn Lys Ser Ile Arg Glu Pro Phe Pro Asp Val His Lys
450 455 460

acc ggc gga gga gaa tcg aaa tgg ttt tac ttc tat ggg ttc ata gct 1440
Thr Gly Gly Gly Glu Ser Lys Trp Phe Tyr Phe Tyr Gly Phe Ile Ala
465 470 475 480

gca ttt ttt gtc gtt gaa gtt tcc ttc att tcg ttt gcg tgg ttc ttt 1488
Ala Phe Phe Val Val Glu Val Ser Phe Ile Ser Phe Ala Trp Phe Phe
485 490 495

gtt ttg aag aga gaa ctc agg cca tct gaa cta tgg gcg tct gag aaa 1536
Val Leu Lys Arg Glu Leu Arg Pro Ser Glu Leu Trp Ala Ser Glu Lys
500 505 510

ggt tac aaa gca atg act agt aat ttt aga agg tac agc tac agg gaa 1584
Gly Tyr Lys Ala Met Thr Ser Asn Phe Arg Arg Tyr Ser Tyr Arg Glu
515 520 525

ctt gtg aag gcg acc aga aaa ttc aag gtt gag cta ggg agg gga gaa 1632
Leu Val Lys Ala Thr Arg Lys Phe Lys Val Glu Leu Gly Arg Gly Glu
530 535 540

tca ggc act gtg tac aaa ggt gtc cta gaa gat gat agg cat gtg gct 1680
Ser Gly Thr Val Tyr Lys Gly Val Leu Glu Asp Asp Arg His Val Ala
545 550 555 560

gtg aag aag ctg gag aat gta agg caa ggc aag gaa gtg ttt cag gct 1728
Val Lys Lys Leu Glu Asn Val Arg Gln Gly Lys Glu Val Phe Gln Ala
565 570 575

gag cta agt gta att ggg agg atc aac cac atg aac ctt gtg agg ata 1776
Glu Leu Ser Val Ile Gly Arg Ile Asn His Met Asn Leu Val Arg Ile
580 585 590

tgg ggc ttc tgt tca gag gga tct cat agg ttg ttg gtc tcc gaa tat 1824
Trp Gly Phe Cys Ser Glu Gly Ser His Arg Leu Leu Val Ser Glu Tyr
595 600 605

gtt gag aat gga tca ctg gct aac att ttg ttc agt gaa gga ggc aac 1872
Val Glu Asn Gly Ser Leu Ala Asn Ile Leu Phe Ser Glu Gly Gly Asn
610 615 620

atc tta ttg gac tgg gag gga agg ttc aac att gcg tta ggt gtt gca 1920
Ile Leu Leu Asp Trp Glu Gly Arg Phe Asn Ile Ala Leu Gly Val Ala

625 630 635 640

aaa ggg tta gcc tat ctc cac cat gag tgc tta gag tgg gtc atc cac 1968
Lys Gly Leu Ala Tyr Leu His His Glu Cys Leu Glu Trp Val Ile His
645 650 655

tgt gat gtg aaa cct gag aac ata ctg tta gac caa gct ttt gag ccc 2016
Cys Asp Val Lys Pro Glu Asn Ile Leu Leu Asp Gln Ala Phe Glu Pro
660 665 670

aag atc act gac ttt ggg ttg gtg aag ttg ctg aac aga gga ggg tcc 2064
Lys Ile Thr Asp Phe Gly Leu Val Lys Leu Leu Asn Arg Gly Gly Ser
675 680 685

acc cag aac gta tcc cat gtc aga gga acg cta ggt tac att gca cct 2112
Thr Gln Asn Val Ser His Val Arg Gly Thr Leu Gly Tyr Ile Ala Pro
690 695 700

gag tgg gtt tcc agc ctc ccg atc aca gca aaa gtc gat gta tac agt 2160
Glu Trp Val Ser Ser Leu Pro Ile Thr Ala Lys Val Asp Val Tyr Ser
705 710 715 720

tat ggg gtt gtg cta ctg gag cta ttg aca ggc acc aga gtt tca gag 2208
Tyr Gly Val Val Leu Leu Glu Leu Leu Thr Gly Thr Arg Val Ser Glu
725 730 735

ttg gtg gga ggc aca gat gag gtg cat agt atg ctt aga aag ctt gtc 2256
Leu Val Gly Gly Thr Asp Glu Val His Ser Met Leu Arg Lys Leu Val
740 745 750

agg atg ctt tct gcc aaa ctt gaa ggg gag gaa caa tcg tgg att gat 2304
Arg Met Leu Ser Ala Lys Leu Glu Gly Glu Glu Gln Ser Trp Ile Asp
755 760 765

ggg tat ctg gat tca aaa ctg aat cgt cca gtc aac tat gtg caa gca 2352
Gly Tyr Leu Asp Ser Lys Leu Asn Arg Pro Val Asn Tyr Val Gln Ala
770 775 780

aga aca ctg atc aaa ttg gcg gtc tcc tgc ttg gag gaa gac aga agc 2400
Arg Thr Leu Ile Lys Leu Ala Val Ser Cys Leu Glu Glu Asp Arg Ser
785 790 795 800

aaa aga ccg act atg gaa cat gca gtc cag acc ctc ctg tca gct gat 2448
Lys Arg Pro Thr Met Glu His Ala Val Gln Thr Leu Leu Ser Ala Asp
805 810 815

gac 2451
Asp

<210> 20
<211> 817
<212> PRT
<213> Zea mays

<400> 20

Met Pro Arg Pro Leu Ala Ala Leu Leu Ser Thr Ala Cys Ile Leu Ser
1 5 10 15

Phe Phe Ile Ala Leu Phe Pro Arg Ala Ala Ser Ser Arg Asp Ile Leu
20 25 30

Pro Leu Gly Ser Ser Leu Val Val Glu Ser Tyr Glu Ser Ser Thr Leu
35 40 45

Gln Ser Ser Asp Gly Thr Phe Ser Ser Gly Phe Tyr Glu Val Tyr Thr
50 55 60

His Ala Phe Thr Phe Ser Val Trp Tyr Ser Lys Thr Glu Ala Ala Ala
65 70 75 80

Ala Asn Asn Lys Thr Ile Val Trp Ser Ala Asn Pro Asp Arg Pro Val
85 90 95

His Ala Arg Arg Ser Ala Leu Thr Leu Gln Lys Asp Gly Asn Met Val
100 105 110

Leu Thr Asp Tyr Asp Gly Ala Ala Val Trp Arg Ala Asp Gly Asn Asn
115 120 125

Phe Thr Gly Val Gln Arg Ala Arg Leu Leu Asp Thr Gly Asn Leu Val
130 135 140

Ile Glu Asp Ser Gly Gly Asn Thr Val Trp Gln Ser Phe Asp Ser Pro
145 150 155 160

Thr Asp Thr Phe Leu Pro Thr Gln Leu Ile Thr Ala Ala Thr Arg Leu
165 170 175

Val Pro Thr Thr Gln Ser Arg Ser Pro Gly Asn Tyr Ile Phe Arg Phe
180 185 190

Ser Asp Leu Ser Val Leu Ser Leu Ile Tyr His Val Pro Gln Val Ser
195 200 205

Asp Ile Tyr Trp Pro Asp Pro Asp Gln Asn Leu Tyr Gln Asp Gly Arg
210 215 220

Asn Gln Tyr Asn Ser Thr Arg Leu Gly Met Leu Thr Asp Ser Gly Val
225 230 235 240

Leu Ala Ser Ser Asp Phe Ala Asp Gly Gln Ala Leu Val Ala Ser Asp
245 250 255

Val Gly Pro Gly Val Lys Arg Arg Leu Thr Leu Asp Pro Asp Gly Asn
260 265 270

Leu Arg Leu Tyr Ser Met Asn Asp Ser Asp Gly Ser Trp Ser Val Ser
275 280 285

Met Val Ala Met Thr Gln Pro Cys Asn Ile His Gly Leu Cys Gly Pro
290 295 300

Asn Gly Ile Cys His Tyr Ser Pro Thr Pro Thr Cys Ser Cys Pro Pro
305 310 315 320

Gly Tyr Ala Thr Arg Asn Pro Gly Asn Trp Thr Glu Gly Cys Met Ala
325 330 335

Ile Val Asn Thr Thr Cys Asp Arg Tyr Asp Lys Arg Ser Met Arg Phe
340 345 350

Val Arg Leu Pro Asn Thr Asp Phe Trp Gly Ser Asp Gln Gln His Leu
355 360 365

Leu Ser Val Ser Leu Arg Thr Cys Arg Asp Ile Cys Ile Ser Asp Cys
370 375 380

Thr Cys Lys Gly Phe Gln Tyr Gln Glu Gly Thr Gly Ser Cys Tyr Pro
385 390 395 400

Lys Ala Tyr Leu Phe Ser Gly Arg Thr Tyr Pro Thr Ser Asp Val Arg
405 410 415

Thr Ile Tyr Leu Lys Leu Pro Thr Gly Val Ser Val Ser Asn Ala Leu
420 425 430

Ile Pro Arg Ser Asp Val Phe Asp Ser Val Pro Arg Arg Leu Asp Cys
435 440 445

Asp Arg Met Asn Lys Ser Ile Arg Glu Pro Phe Pro Asp Val His Lys
450 455 460

Thr Gly Gly Gly Glu Ser Lys Trp Phe Tyr Phe Tyr Gly Phe Ile Ala
465 470 475 480

Ala Phe Phe Val Val Glu Val Ser Phe Ile Ser Phe Ala Trp Phe Phe
485 490 495

Val Leu Lys Arg Glu Leu Arg Pro Ser Glu Leu Trp Ala Ser Glu Lys

500 505 510

Gly Tyr Lys Ala Met Thr Ser Asn Phe Arg Arg Tyr Ser Tyr Arg Glu
515 520 525

Leu Val Lys Ala Thr Arg Lys Phe Lys Val Glu Leu Gly Arg Gly Glu
530 535 540

Ser Gly Thr Val Tyr Lys Gly Val Leu Glu Asp Asp Arg His Val Ala
545 550 555 560

Val Lys Lys Leu Glu Asn Val Arg Gln Gly Lys Glu Val Phe Gln Ala
565 570 575

Glu Leu Ser Val Ile Gly Arg Ile Asn His Met Asn Leu Val Arg Ile
580 585 590

Trp Gly Phe Cys Ser Glu Gly Ser His Arg Leu Leu Val Ser Glu Tyr
595 600 605

Val Glu Asn Gly Ser Leu Ala Asn Ile Leu Phe Ser Glu Gly Gly Asn
610 615 620

Ile Leu Leu Asp Trp Glu Gly Arg Phe Asn Ile Ala Leu Gly Val Ala
625 630 635 640

Lys Gly Leu Ala Tyr Leu His His Glu Cys Leu Glu Trp Val Ile His
645 650 655

Cys Asp Val Lys Pro Glu Asn Ile Leu Leu Asp Gln Ala Phe Glu Pro
660 665 670

Lys Ile Thr Asp Phe Gly Leu Val Lys Leu Leu Asn Arg Gly Gly Ser
675 680 685

Thr Gln Asn Val Ser His Val Arg Gly Thr Leu Gly Tyr Ile Ala Pro
690 695 700

Glu Trp Val Ser Ser Leu Pro Ile Thr Ala Lys Val Asp Val Tyr Ser
705 710 715 720

Tyr Gly Val Val Leu Leu Glu Leu Leu Thr Gly Thr Arg Val Ser Glu
725 730 735

Leu Val Gly Gly Thr Asp Glu Val His Ser Met Leu Arg Lys Leu Val
740 745 750

Arg Met Leu Ser Ala Lys Leu Glu Gly Glu Glu Gln Ser Trp Ile Asp
755 760 765

Gly Tyr Leu Asp Ser Lys Leu Asn Arg Pro Val Asn Tyr Val Gln Ala
770 775 780

Arg Thr Leu Ile Lys Leu Ala Val Ser Cys Leu Glu Glu Asp Arg Ser
785 790 795 800

Lys Arg Pro Thr Met Glu His Ala Val Gln Thr Leu Leu Ser Ala Asp
805 810 815

Asp

<210> 21

<211> 1434

<212> DNA

<213> Zea mays

<220>

<221> CDS

<222> (1)..(1434)

<223> RECEPTOR PROTEIN KINASE ZMPK1 PRECURSOR maize

<400> 21
atg ggg atc agc aag gga ggc agc ggc aag gag gcg aag aag ccg ccg 48
Met Gly Ile Ser Lys Gly Gly Ser Gly Lys Glu Ala Lys Lys Pro Pro
1 5 10 15

ctg ctg ctg ggg cga ttc gag gtc ggg aag ctg ctg ggg cag ggc aac 96
Leu Leu Leu Gly Arg Phe Glu Val Gly Lys Leu Leu Gly Gln Gly Asn
20 25 30

ttc gcc aag gtg tac cac gcg cgc aac gtg gcc acc ggc gag gag gtg 144
Phe Ala Lys Val Tyr His Ala Arg Asn Val Ala Thr Gly Glu Glu Val
35 40 45

gcg atc aag gtg atg gag aag gag aag atc ttc aag tcg ggg ctc acg 192
Ala Ile Lys Val Met Glu Lys Glu Lys Ile Phe Lys Ser Gly Leu Thr
50 55 60

```
gcg cac atc aag cgg gag atc gcc gtg ctc cgg cgc gtc cgc cac ccg      240
Ala His Ile Lys Arg Glu Ile Ala Val Leu Arg Arg Val Arg His Pro
65                  70                  75                  80

cac atc gtg cag ctg tac gag gtg atg gcc acc aag ctc cgg atc tac      288
His Ile Val Gln Leu Tyr Glu Val Met Ala Thr Lys Leu Arg Ile Tyr
                85                  90                  95

ttc gtc atg gag tac gtc cgc ggc ggc gag ctg ttc gcg cgc gtg gcg      336
Phe Val Met Glu Tyr Val Arg Gly Gly Glu Leu Phe Ala Arg Val Ala
            100                 105                 110

cgg ggg cgg ctg ccc gag gcc gac gcg cgg cgc tac ttc cag cag ctg      384
Arg Gly Arg Leu Pro Glu Ala Asp Ala Arg Arg Tyr Phe Gln Gln Leu
        115                 120                 125

gtg tcc gcc gtc gcg ttc tgc cac gcg cgc ggg gtg ttc cac cgc gac      432
Val Ser Ala Val Ala Phe Cys His Ala Arg Gly Val Phe His Arg Asp
    130                 135                 140

atc aag ccg gag aac ctc ctc gtc gac gac gcc ggc gac ctc aag gtg      480
Ile Lys Pro Glu Asn Leu Leu Val Asp Asp Ala Gly Asp Leu Lys Val
145                 150                 155                 160

tcc gac ttc ggg ctc tcc gcg gtg gcg gac ggg atg cgg cgc gac ggg      528
Ser Asp Phe Gly Leu Ser Ala Val Ala Asp Gly Met Arg Arg Asp Gly
                165                 170                 175

ctg ttc cac acg ttc tgc ggc acg ccg gcg tac gtc gcg ccg gag gtg      576
Leu Phe His Thr Phe Cys Gly Thr Pro Ala Tyr Val Ala Pro Glu Val
            180                 185                 190

ctg tcg cgc cgc ggg tac gac gcc gcc ggg gcc gac ctc tgg tcc tgc      624
Leu Ser Arg Arg Gly Tyr Asp Ala Ala Gly Ala Asp Leu Trp Ser Cys
        195                 200                 205

ggc gtc gtg ctc ttc gtc ctc atg gcc ggc tac ctc ccc ttc cag gac      672
Gly Val Val Leu Phe Val Leu Met Ala Gly Tyr Leu Pro Phe Gln Asp
    210                 215                 220

cgc aac ctc gcc ggc atg tac cgc aag atc cac aag ggc gac ttc cgc      720
Arg Asn Leu Ala Gly Met Tyr Arg Lys Ile His Lys Gly Asp Phe Arg
225                 230                 235                 240

tgc ccc aag tgg ttc tcg ccg gag ctc atc cgc ctc ctc cgc ggc gtc      768
Cys Pro Lys Trp Phe Ser Pro Glu Leu Ile Arg Leu Leu Arg Gly Val
                245                 250                 255

ctc gtc acc aac ccg cag cgc cgc gcc acc gcc gag ggg atc atg gag      816
Leu Val Thr Asn Pro Gln Arg Arg Ala Thr Ala Glu Gly Ile Met Glu
            260                 265                 270

aac gag tgg ttc aag atc ggc ttc cgc cgc ttc tcc ttc cgc gtc gag      864
Asn Glu Trp Phe Lys Ile Gly Phe Arg Arg Phe Ser Phe Arg Val Glu
        275                 280                 285

gac gac cgc acc ttc acc tgc ttc gaa ctt gac gac gac gcc gcc gtc      912
Asp Asp Arg Thr Phe Thr Cys Phe Glu Leu Asp Asp Asp Ala Ala Val
    290                 295                 300

gac gcg ccc acc tcg ccg ccg gac acg ccg cgg aca gtg gac agc ggc      960
```

Asp Ala Pro Thr Ser Pro Pro Asp Thr Pro Arg Thr Val Asp Ser Gly
305 310 315 320

gac gtc ggc gct gct ccg acg cga cca aga aaa gcc ggg agc ctg acg 1008
Asp Val Gly Ala Ala Pro Thr Arg Pro Arg Lys Ala Gly Ser Leu Thr
325 330 335

tcg tgc gac tcg gcg ccc ctg aac gcg ttc gac atc atc tcc ttc tcc 1056
Ser Cys Asp Ser Ala Pro Leu Asn Ala Phe Asp Ile Ile Ser Phe Ser
340 345 350

ccg ggg ttc gac ctc tca gga ctc atc ccg gag cag cag aaa cac acg 1104
Pro Gly Phe Asp Leu Ser Gly Leu Ile Pro Glu Gln Gln Lys His Thr
355 360 365

gcg agg ttc gtg tcg gcg gcg ccg gtg gag gtg atc gtg gcg acg ctg 1152
Ala Arg Phe Val Ser Ala Ala Pro Val Glu Val Ile Val Ala Thr Leu
370 375 380

gag gcg gcc gcg gcg gcg gcg ggc atg gcg gtg cgg gag agg gag gac 1200
Glu Ala Ala Ala Ala Ala Ala Gly Met Ala Val Arg Glu Arg Glu Asp
385 390 395 400

ggg tcg atc agc atg gag ggg aca cgc gag ggc gag cac ggc gcg ctg 1248
Gly Ser Ile Ser Met Glu Gly Thr Arg Glu Gly Glu His Gly Ala Leu
405 410 415

gcg gtg gcc gcg gag atc tac gag ctc acg ccg gag ctg gtg gtg gtg 1296
Ala Val Ala Ala Glu Ile Tyr Glu Leu Thr Pro Glu Leu Val Val Val
420 425 430

gag gtg cgg cgg aag gcc ggc ggc gcc gcc gag tac gag gag ttc ttc 1344
Glu Val Arg Arg Lys Ala Gly Gly Ala Ala Glu Tyr Glu Glu Phe Phe
435 440 445

cgg gcg cgg ctc aag cca agc ctc cgc gag ctc gtc tgc gac gac cgg 1392
Arg Ala Arg Leu Lys Pro Ser Leu Arg Glu Leu Val Cys Asp Asp Arg
450 455 460

cca tgc ccg gag gac tcc ggc gag ctc tcc cgg agc ctt tga 1434
Pro Cys Pro Glu Asp Ser Gly Glu Leu Ser Arg Ser Leu
465 470 475

<210> 22

<211> 477

<212> PRT

<213> Zea mays

<400> 22

Met Gly Ile Ser Lys Gly Gly Ser Gly Lys Glu Ala Lys Lys Pro Pro
1 5 10 15

Leu Leu Leu Gly Arg Phe Glu Val Gly Lys Leu Leu Gly Gln Gly Asn

20 25 30

Phe Ala Lys Val Tyr His Ala Arg Asn Val Ala Thr Gly Glu Glu Val
35 40 45

Ala Ile Lys Val Met Glu Lys Glu Lys Ile Phe Lys Ser Gly Leu Thr
50 55 60

Ala His Ile Lys Arg Glu Ile Ala Val Leu Arg Arg Val Arg His Pro
65 70 75 80

His Ile Val Gln Leu Tyr Glu Val Met Ala Thr Lys Leu Arg Ile Tyr
85 90 95

Phe Val Met Glu Tyr Val Arg Gly Gly Glu Leu Phe Ala Arg Val Ala
100 105 110

Arg Gly Arg Leu Pro Glu Ala Asp Ala Arg Arg Tyr Phe Gln Gln Leu
115 120 125

Val Ser Ala Val Ala Phe Cys His Ala Arg Gly Val Phe His Arg Asp
130 135 140

Ile Lys Pro Glu Asn Leu Leu Val Asp Asp Ala Gly Asp Leu Lys Val
145 150 155 160

Ser Asp Phe Gly Leu Ser Ala Val Ala Asp Gly Met Arg Arg Asp Gly
165 170 175

Leu Phe His Thr Phe Cys Gly Thr Pro Ala Tyr Val Ala Pro Glu Val
180 185 190

Leu Ser Arg Arg Gly Tyr Asp Ala Ala Gly Ala Asp Leu Trp Ser Cys
195 200 205

Gly Val Val Leu Phe Val Leu Met Ala Gly Tyr Leu Pro Phe Gln Asp
210 215 220

Arg Asn Leu Ala Gly Met Tyr Arg Lys Ile His Lys Gly Asp Phe Arg
225 230 235 240

Cys Pro Lys Trp Phe Ser Pro Glu Leu Ile Arg Leu Leu Arg Gly Val
245 250 255

Leu Val Thr Asn Pro Gln Arg Arg Ala Thr Ala Glu Gly Ile Met Glu
260 265 270

Asn Glu Trp Phe Lys Ile Gly Phe Arg Arg Phe Ser Phe Arg Val Glu
275 280 285

Asp Asp Arg Thr Phe Thr Cys Phe Glu Leu Asp Asp Asp Ala Ala Val
290 295 300

Asp Ala Pro Thr Ser Pro Pro Asp Thr Pro Arg Thr Val Asp Ser Gly
305 310 315 320

Asp Val Gly Ala Ala Pro Thr Arg Pro Arg Lys Ala Gly Ser Leu Thr
325 330 335

Ser Cys Asp Ser Ala Pro Leu Asn Ala Phe Asp Ile Ile Ser Phe Ser
340 345 350

Pro Gly Phe Asp Leu Ser Gly Leu Ile Pro Glu Gln Gln Lys His Thr
355 360 365

Ala Arg Phe Val Ser Ala Ala Pro Val Glu Val Ile Val Ala Thr Leu
370 375 380

Glu Ala Ala Ala Ala Ala Ala Gly Met Ala Val Arg Glu Arg Glu Asp
385 390 395 400

Gly Ser Ile Ser Met Glu Gly Thr Arg Glu Gly Glu His Gly Ala Leu
405 410 415

Ala Val Ala Ala Glu Ile Tyr Glu Leu Thr Pro Glu Leu Val Val Val
420 425 430

Glu Val Arg Arg Lys Ala Gly Gly Ala Ala Glu Tyr Glu Glu Phe Phe
435 440 445

Arg Ala Arg Leu Lys Pro Ser Leu Arg Glu Leu Val Cys Asp Asp Arg
450 455 460

Pro Cys Pro Glu Asp Ser Gly Glu Leu Ser Arg Ser Leu
465 470 475

<210> 23
<211> 672
<212> DNA
<213> Zea mays

<220>
<221> CDS

<222> (1)..(672)
<223> PIR2:T04358 Glutathione S-Transferase 1
glutathione transferase (EC 2.5.1.18)

<400> 23
```
atg gcc gag gag aag aag cag ggc ctg cag ctg ctg gac ttc tgg gtg      48
Met Ala Glu Glu Lys Lys Gln Gly Leu Gln Leu Leu Asp Phe Trp Val
1               5                   10                  15

agc cca ttc ggg cag cgc tgc cgc atc gcg atg gac gag aag ggc ctg      96
Ser Pro Phe Gly Gln Arg Cys Arg Ile Ala Met Asp Glu Lys Gly Leu
            20                  25                  30

gcc tac gag tac ctg gag cag gac ctg ggg aac aag agc gag ctg ctc     144
Ala Tyr Glu Tyr Leu Glu Gln Asp Leu Gly Asn Lys Ser Glu Leu Leu
        35                  40                  45

ctc cgc gcc aac ccg gtg cat aag aag atc ccc gtg ctg ctg cac gac     192
Leu Arg Ala Asn Pro Val His Lys Lys Ile Pro Val Leu Leu His Asp
    50                  55                  60

ggc cgc ccc gtc tgc gag tcc ctc gtc atc gtg cag tac ctc gac gag     240
Gly Arg Pro Val Cys Glu Ser Leu Val Ile Val Gln Tyr Leu Asp Glu
65                  70                  75                  80

gcg ttc ccg gcg gcg gcg ccg gcg ctg ctc ccc gcc gac ccc tac gcg     288
Ala Phe Pro Ala Ala Ala Pro Ala Leu Leu Pro Ala Asp Pro Tyr Ala
                85                  90                  95

cgc gcg cag gcc cgc ttc tgg gcg gac tac gtc gac aag aag ctc tac     336
Arg Ala Gln Ala Arg Phe Trp Ala Asp Tyr Val Asp Lys Lys Leu Tyr
            100                 105                 110

gac tgc ggc acc cgg ctg tgg aag ctc aag ggg gac ggc cag gcg cag     384
Asp Cys Gly Thr Arg Leu Trp Lys Leu Lys Gly Asp Gly Gln Ala Gln
        115                 120                 125

gcg cgc gcc gag atg gtc gag atc ctc cgc acg ctg gag ggc gcg ctc     432
Ala Arg Ala Glu Met Val Glu Ile Leu Arg Thr Leu Glu Gly Ala Leu
    130                 135                 140

ggc gac ggg ccc ttc ttc ggc ggc gac gcc ctc ggc ttc gtc gac gtc     480
Gly Asp Gly Pro Phe Phe Gly Gly Asp Ala Leu Gly Phe Val Asp Val
145                 150                 155                 160

gcg ctc gtg ccc ttc acg tcc tgg ttc ctc gcc tac gac cgc ttc ggc     528
Ala Leu Val Pro Phe Thr Ser Trp Phe Leu Ala Tyr Asp Arg Phe Gly
                165                 170                 175

ggc gtc agc gtg gag aag gag tgc ccg agg ctg gcc gcc tgg gcc aag     576
Gly Val Ser Val Glu Lys Glu Cys Pro Arg Leu Ala Ala Trp Ala Lys
            180                 185                 190

cgc tgc gcc gag cgc ccc agc gtc gcc aag aac ctc tac ccg ccc gag     624
Arg Cys Ala Glu Arg Pro Ser Val Ala Lys Asn Leu Tyr Pro Pro Glu
        195                 200                 205

aag gtc tac gac ttc gtc tgc ggg atg aag aag agg ctg ggc atc gag     672
Lys Val Tyr Asp Phe Val Cys Gly Met Lys Lys Arg Leu Gly Ile Glu
    210                 215                 220
```

<210> 24
<211> 224
<212> PRT
<213> Zea mays

<400> 24

Met Ala Glu Glu Lys Lys Gln Gly Leu Gln Leu Leu Asp Phe Trp Val
1   5   10   15

Ser Pro Phe Gly Gln Arg Cys Arg Ile Ala Met Asp Glu Lys Gly Leu
            20   25   30

Ala Tyr Glu Tyr Leu Glu Gln Asp Leu Gly Asn Lys Ser Glu Leu Leu
        35   40   45

Leu Arg Ala Asn Pro Val His Lys Lys Ile Pro Val Leu Leu His Asp
    50   55   60

Gly Arg Pro Val Cys Glu Ser Leu Val Ile Val Gln Tyr Leu Asp Glu
65   70   75   80

Ala Phe Pro Ala Ala Ala Pro Ala Leu Leu Pro Ala Asp Pro Tyr Ala
                85   90   95

Arg Ala Gln Ala Arg Phe Trp Ala Asp Tyr Val Asp Lys Lys Leu Tyr
            100   105   110

Asp Cys Gly Thr Arg Leu Trp Lys Leu Lys Gly Asp Gly Gln Ala Gln
        115   120   125

Ala Arg Ala Glu Met Val Glu Ile Leu Arg Thr Leu Glu Gly Ala Leu
    130   135   140

Gly Asp Gly Pro Phe Phe Gly Gly Asp Ala Leu Gly Phe Val Asp Val
145   150   155   160

Ala Leu Val Pro Phe Thr Ser Trp Phe Leu Ala Tyr Asp Arg Phe Gly
                165   170   175

Gly Val Ser Val Glu Lys Glu Cys Pro Arg Leu Ala Ala Trp Ala Lys
            180   185   190

Arg Cys Ala Glu Arg Pro Ser Val Ala Lys Asn Leu Tyr Pro Pro Glu
        195   200   205

```
Lys Val Tyr Asp Phe Val Cys Gly Met Lys Lys Arg Leu Gly Ile Glu
    210                 215                 220


<210>  25

<211>  945

<212>  DNA

<213>  Oryza sativa


<220>

<221>  CDS

<222>  (6)..(698)

<223>  Glutathione S-Transferase 1


<400>  25
tagcc atg gcg gag gag aag gag ctg gtg ctg ctc gat ttc tgg gtg agc      50
      Met Ala Glu Glu Lys Glu Leu Val Leu Leu Asp Phe Trp Val Ser
      1               5                   10                  15

ccg ttc ggg cag cgg tgc cgg atc gcc atg gcg gag aag ggg ctg gag        98
Pro Phe Gly Gln Arg Cys Arg Ile Ala Met Ala Glu Lys Gly Leu Glu
                20                  25                  30

ttc gag tac cgc gag gag gac ctc ggc aac aag agc gac ctc ctc ctc       146
Phe Glu Tyr Arg Glu Glu Asp Leu Gly Asn Lys Ser Asp Leu Leu Leu
            35                  40                  45

cgc tcc aac ccc gtc cac agg aag atc ccc gtc ctc ctc cac gcc ggc       194
Arg Ser Asn Pro Val His Arg Lys Ile Pro Val Leu Leu His Ala Gly
        50                  55                  60

cgc ccc gtc tcc gag tcc ctc gtc atc ctc cag tac ctc gac gac gcg       242
Arg Pro Val Ser Glu Ser Leu Val Ile Leu Gln Tyr Leu Asp Asp Ala
    65                  70                  75

ttc ccc ggc acc ccc cac ctc ctc cct ccg ggg aac tcc ggc gac gcc       290
Phe Pro Gly Thr Pro His Leu Leu Pro Pro Gly Asn Ser Gly Asp Ala
80                  85                  90                  95

gac gcc gcg ttc gcg cgc gcc acg gcg agg ttc tgg gcg gac tac gtc       338
Asp Ala Ala Phe Ala Arg Ala Thr Ala Arg Phe Trp Ala Asp Tyr Val
                100                 105                 110

gac agg aag ctc tac gac tgc ggg tcc agg ctg tgg agg ctc aag ggt       386
Asp Arg Lys Leu Tyr Asp Cys Gly Ser Arg Leu Trp Arg Leu Lys Gly
            115                 120                 125

gag ccg cat gcg gcg gcg ggg cgc gag atg gcg gag atc ctc cgc acg       434
Glu Pro His Ala Ala Ala Gly Arg Glu Met Ala Glu Ile Leu Arg Thr
        130                 135                 140
```

ctg gag gcg gag ctc ggc gac cgg gag ttc ttc ggc ggc ggc ggc ggc 482
Leu Glu Ala Glu Leu Gly Asp Arg Glu Phe Phe Gly Gly Gly Gly Gly
145 150 155

ggc agg ctc ggg ttc gtc gac gtc gcg ctc gtg ccg ttc acg gcg tgt 530
Gly Arg Leu Gly Phe Val Asp Val Ala Leu Val Pro Phe Thr Ala Cys
160 165 170 175

tcc aca gct act gag agg tgc ggc ggg ttc agc gtg gag gag gtg gcg 578
Ser Thr Ala Thr Glu Arg Cys Gly Gly Phe Ser Val Glu Glu Val Ala
180 185 190

ccg agg ctg gcg gcg tgg gcg cgg cgg cgc ggc cgg atc gac tcc gtc 626
Pro Arg Leu Ala Ala Trp Ala Arg Arg Arg Gly Arg Ile Asp Ser Val
195 200 205

gtc aag cac ctc ccc tcg ccg gag aag gtc tac gac ttc gtc ggc gtc 674
Val Lys His Leu Pro Ser Pro Glu Lys Val Tyr Asp Phe Val Gly Val
210 215 220

ctc aag aag aag tac ggc gtc gag tagatcggtg gatgcgaagt tgcagggatc 728
Leu Lys Lys Lys Tyr Gly Val Glu
225 230

gattggcggt tgcgttcgca acgtgaacga ttcgtccgtt gtttcagtgg ccaagtgtgt 788

gtgagtttgt tgttaccgtt gagtgcttgt gtgtgggatg gttggtggca gcagagagtt 848

gcctccgatt ctctgagata gtcactaaat aaagtttgtc ctttgaaact aaaaaaagtt 908

ggctttggtt aaaaaaaaaa aaaaaaaaaa aaaaaaa 945

<210> 26

<211> 231

<212> PRT

<213> Oryza sativa

<400> 26

Met Ala Glu Glu Lys Glu Leu Val Leu Leu Asp Phe Trp Val Ser Pro
1 5 10 15

Phe Gly Gln Arg Cys Arg Ile Ala Met Ala Glu Lys Gly Leu Glu Phe
20 25 30

Glu Tyr Arg Glu Glu Asp Leu Gly Asn Lys Ser Asp Leu Leu Leu Arg
35 40 45

Ser Asn Pro Val His Arg Lys Ile Pro Val Leu Leu His Ala Gly Arg
50 55 60

Pro Val Ser Glu Ser Leu Val Ile Leu Gln Tyr Leu Asp Asp Ala Phe
65 70 75 80

Pro Gly Thr Pro His Leu Leu Pro Pro Gly Asn Ser Gly Asp Ala Asp
85 90 95

Ala Ala Phe Ala Arg Ala Thr Ala Arg Phe Trp Ala Asp Tyr Val Asp
100 105 110

Arg Lys Leu Tyr Asp Cys Gly Ser Arg Leu Trp Arg Leu Lys Gly Glu
115 120 125

Pro His Ala Ala Ala Gly Arg Glu Met Ala Glu Ile Leu Arg Thr Leu
130 135 140

Glu Ala Glu Leu Gly Asp Arg Glu Phe Phe Gly Gly Gly Gly Gly Gly
145 150 155 160

Arg Leu Gly Phe Val Asp Val Ala Leu Val Pro Phe Thr Ala Cys Ser
165 170 175

Thr Ala Thr Glu Arg Cys Gly Gly Phe Ser Val Glu Glu Val Ala Pro
180 185 190

Arg Leu Ala Ala Trp Ala Arg Arg Arg Gly Arg Ile Asp Ser Val Val
195 200 205

Lys His Leu Pro Ser Pro Glu Lys Val Tyr Asp Phe Val Gly Val Leu
210 215 220

Lys Lys Lys Tyr Gly Val Glu
225 230

<210> 27
<211> 486
<212> DNA
<213> Oryza sativa

<220>
<221> CDS
<222> (1)..(486)
<223> LOCUS AF203879 peroxiredoxin ACCESSION AF203879
Thioredoxin-dependent peroxidase

<400> 27
atg gcc ccg gtt gcc gtg ggc gac acc ctc ccc gac ggc cag ctg ggg 48
Met Ala Pro Val Ala Val Gly Asp Thr Leu Pro Asp Gly Gln Leu Gly

```
1               5                   10                  15

tgg ttc gac ggg gag gac aag ctg cag cag gtc tcc gtc cac ggc ctc        96
Trp Phe Asp Gly Glu Asp Lys Leu Gln Gln Val Ser Val His Gly Leu
            20                  25                  30

gcc gcc ggc aag aag gtc gtc ctc ttc ggc gtc ccc ggt gcc ttc acc       144
Ala Ala Gly Lys Lys Val Val Leu Phe Gly Val Pro Gly Ala Phe Thr
        35                  40                  45

ccg acc tgc agc aat cag cat gtg cca gga ttc ata aat cag gct gag       192
Pro Thr Cys Ser Asn Gln His Val Pro Gly Phe Ile Asn Gln Ala Glu
    50                  55                  60

cag ctc aaa gcc aag ggt gta gac gac atc ttg ctt gtc agt gtt aac       240
Gln Leu Lys Ala Lys Gly Val Asp Asp Ile Leu Leu Val Ser Val Asn
65                  70                  75                  80

gac ccc ttt gtc atg aag gcg tgg gca aag tca tac cct gag aat aag       288
Asp Pro Phe Val Met Lys Ala Trp Ala Lys Ser Tyr Pro Glu Asn Lys
                85                  90                  95

cat gtg aaa ttc ctt gcc gat ggt ttg gga aca tac acc aag gca ctt       336
His Val Lys Phe Leu Ala Asp Gly Leu Gly Thr Tyr Thr Lys Ala Leu
            100                 105                 110

ggt ctt gag ctt gac ctt tcg gag aaa ggg ctt ggt att cgt tcg aga       384
Gly Leu Glu Leu Asp Leu Ser Glu Lys Gly Leu Gly Ile Arg Ser Arg
        115                 120                 125

cgg ttt gct ctc ctt gct gac aac ctc aag gtt act gtt gca aac att       432
Arg Phe Ala Leu Leu Ala Asp Asn Leu Lys Val Thr Val Ala Asn Ile
    130                 135                 140

gag gaa ggt ggc caa ttc aca atc tct ggt gct gag gag atc ctc aag       480
Glu Glu Gly Gly Gln Phe Thr Ile Ser Gly Ala Glu Glu Ile Leu Lys
145                 150                 155                 160

gca ctg                                                               486
Ala Leu
```

<210> 28
<211> 162
<212> PRT
<213> Oryza sativa

<400> 28

```
Met Ala Pro Val Ala Val Gly Asp Thr Leu Pro Asp Gly Gln Leu Gly
1               5                   10                  15

Trp Phe Asp Gly Glu Asp Lys Leu Gln Gln Val Ser Val His Gly Leu
            20                  25                  30

Ala Ala Gly Lys Lys Val Val Leu Phe Gly Val Pro Gly Ala Phe Thr
        35                  40                  45
```

Pro Thr Cys Ser Asn Gln His Val Pro Gly Phe Ile Asn Gln Ala Glu
50 55 60

Gln Leu Lys Ala Lys Gly Val Asp Asp Ile Leu Leu Val Ser Val Asn
65 70 75 80

Asp Pro Phe Val Met Lys Ala Trp Ala Lys Ser Tyr Pro Glu Asn Lys
85 90 95

His Val Lys Phe Leu Ala Asp Gly Leu Gly Thr Tyr Thr Lys Ala Leu
100 105 110

Gly Leu Glu Leu Asp Leu Ser Glu Lys Gly Leu Gly Ile Arg Ser Arg
115 120 125

Arg Phe Ala Leu Leu Ala Asp Asn Leu Lys Val Thr Val Ala Asn Ile
130 135 140

Glu Glu Gly Gly Gln Phe Thr Ile Ser Gly Ala Glu Glu Ile Leu Lys
145 150 155 160

Ala Leu

<210> 29

<211> 647

<212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS

<222> (60)..(548)

<223> Thioredoxin-dependent peroxidase

<400> 29
ccacgcgtcc gcaaaactct tctattttcc tctgtcttca aaaccacaga gatctcttc 59

atg gct cca att act gtc ggc gat gtt gta cca gac gga act atc tct 107
Met Ala Pro Ile Thr Val Gly Asp Val Val Pro Asp Gly Thr Ile Ser
1 5 10 15

ttc ttc gat gaa aat gat cag ctt cag acc gtc tcc gtt cac tct atc 155

```
Phe Phe Asp Glu Asn Asp Gln Leu Gln Thr Val Ser Val His Ser Ile
            20                  25                  30

gcc gcc ggt aaa aaa gtc att ctc ttt ggt gtt cct ggt gct ttc act      203
Ala Ala Gly Lys Lys Val Ile Leu Phe Gly Val Pro Gly Ala Phe Thr
        35                  40                  45

ccc aca tgc agt atg agc cat gtg cct gga ttc att ggg aaa gca gag      251
Pro Thr Cys Ser Met Ser His Val Pro Gly Phe Ile Gly Lys Ala Glu
    50                  55                  60

gag ctg aag tca aag ggt att gat gag atc att tgc ttt agt gtg aat      299
Glu Leu Lys Ser Lys Gly Ile Asp Glu Ile Ile Cys Phe Ser Val Asn
65                  70                  75                  80

gat cca ttt gtg atg aag gca tgg gga aaa aca tat cca gag aac aag      347
Asp Pro Phe Val Met Lys Ala Trp Gly Lys Thr Tyr Pro Glu Asn Lys
                85                  90                  95

cat gtg aag ttt gta gca gat ggg tct gga gaa tac acg cat ctt ctt      395
His Val Lys Phe Val Ala Asp Gly Ser Gly Glu Tyr Thr His Leu Leu
            100                 105                 110

gga ctt gag ctt gac ctt aag gac aag ggt tct ggt att agt tca ggg      443
Gly Leu Glu Leu Asp Leu Lys Asp Lys Gly Ser Gly Ile Ser Ser Gly
        115                 120                 125

aga ttc gct ttg ttg ctt gat aac ctt aag gtg act gta gcc aat gtt      491
Arg Phe Ala Leu Leu Leu Asp Asn Leu Lys Val Thr Val Ala Asn Val
    130                 135                 140

gaa tct ggt ggc gag ttc acg gtt tcc agc gca gag gat att ctc aag      539
Glu Ser Gly Gly Glu Phe Thr Val Ser Ser Ala Glu Asp Ile Leu Lys
145                 150                 155                 160

gct ctt taa gaaactttat cgtttcgctt gttgtattgt gaatctaaac              588
Ala Leu

tgctgtatgt gaagaagaga tttctatagc ttgatttcaa tcaaaaaaaa aaaaaaaaa     647
```

<210> 30

<211> 162

<212> PRT

<213> Arabidopsis thaliana

<400> 30

```
Met Ala Pro Ile Thr Val Gly Asp Val Val Pro Asp Gly Thr Ile Ser
1               5                   10                  15

Phe Phe Asp Glu Asn Asp Gln Leu Gln Thr Val Ser Val His Ser Ile
            20                  25                  30
```

Ala Ala Gly Lys Lys Val Ile Leu Phe Gly Val Pro Gly Ala Phe Thr
35 40 45

Pro Thr Cys Ser Met Ser His Val Pro Gly Phe Ile Gly Lys Ala Glu
50 55 60

Glu Leu Lys Ser Lys Gly Ile Asp Glu Ile Ile Cys Phe Ser Val Asn
65 70 75 80

Asp Pro Phe Val Met Lys Ala Trp Gly Lys Thr Tyr Pro Glu Asn Lys
85 90 95

His Val Lys Phe Val Ala Asp Gly Ser Gly Glu Tyr Thr His Leu Leu
100 105 110

Gly Leu Glu Leu Asp Leu Lys Asp Lys Gly Ser Gly Ile Ser Ser Gly
115 120 125

Arg Phe Ala Leu Leu Leu Asp Asn Leu Lys Val Thr Val Ala Asn Val
130 135 140

Glu Ser Gly Gly Glu Phe Thr Val Ser Ser Ala Glu Asp Ile Leu Lys
145 150 155 160

Ala Leu

<210> 31
<211> 1846
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (494)..(1159)
<223> Rab28 protein; abscisic acid inducible; rab28 gene
EMBL no. X59138;
PIR2 no. S18545

<220>
<221> CDS
<222> (1289)..(1456)
<223>

<400> 31
gtcgacggtg gaaactggaa acgggcctaa tggatgcggt actaagccag gtccagtaaa 60

tttgtcgtgc agagatgccc cgtggacttt agaggagccc acgataggac tctggcgcag 120

```
ttcggcctgc aagtaaaggc ccatgactcg ccgccgccgt ctggaccgct gacacgcatg   180

ccctgatgct ccccttccg ggagcttctt catccagctt gcagccggac agcccttgcg    240

ctcgcgccac gtgggcatgc cgccgcgacg cgcctcctct tgctcgtctc cacgtctctc   300

gcctctgcaa cacatgacat atgtccctcc tcgagctccc cgcacgccgc atataatcgc   360

aacgatccaa tccacactca ccatcaagtc tcaagaggca gctaaattaa accaacaagc   420

cgtgatctgt actgtagtag caacttggtt cttggtaggt gagatagtga tacagcaggt   480

agctcgagcg aga atg agc cag gag cag ccg agg agg ccg tcc ggc cat      529
               Met Ser Gln Glu Gln Pro Arg Arg Pro Ser Gly His
               1               5                   10

gag gag acg agc ggc ggc gga gag cag ggc gcc gtc cgc tac ggc gac     577
Glu Glu Thr Ser Gly Gly Gly Glu Gln Gly Ala Val Arg Tyr Gly Asp
        15                  20                  25

gtg ttc ccg gcg gtg agc ggg ggc ctc gcg gag aag ccc gtg gcg cgc     625
Val Phe Pro Ala Val Ser Gly Gly Leu Ala Glu Lys Pro Val Ala Arg
    30                  35                  40

agg acc gcc acg atg cag tcg gcg gag aac ctg gtg ttc ggc cag acg     673
Arg Thr Ala Thr Met Gln Ser Ala Glu Asn Leu Val Phe Gly Gln Thr
45                  50                  55                  60

ctc aag ggc ggc ccg gcg gcg gcc atg cag tcc gcg gcc acc acc aac     721
Leu Lys Gly Gly Pro Ala Ala Ala Met Gln Ser Ala Ala Thr Thr Asn
                65                  70                  75

gag cgc atg ggc gcc gtc ggg cac gac cag gcc acg gac gcc acc gcc     769
Glu Arg Met Gly Ala Val Gly His Asp Gln Ala Thr Asp Ala Thr Ala
            80                  85                  90

gtg cag ggc gtc acc gtc tcc gag acc cgc gtc cct ggc ggc ggc cgc     817
Val Gln Gly Val Thr Val Ser Glu Thr Arg Val Pro Gly Gly Gly Arg
        95                  100                 105

atc gtc acc gag ttc gtc gcc ggg cag gct gtc ggc cag tac ctc gcg     865
Ile Val Thr Glu Phe Val Ala Gly Gln Ala Val Gly Gln Tyr Leu Ala
    110                 115                 120

cgg gac gac gat ggc ggc ggc ggc atc gcc ggc ccc ggc gcc gga gcg     913
Arg Asp Asp Asp Gly Gly Gly Gly Ile Ala Gly Pro Gly Ala Gly Ala
125                 130                 135                 140

gga gtt gca ggt aag gat atc aca aag gtg acc atc ggc gag gcg ctc     961
Gly Val Ala Gly Lys Asp Ile Thr Lys Val Thr Ile Gly Glu Ala Leu
                145                 150                 155

gag gcg acg gcg ctc gcg gcg ggt gac gcg ccg gtg gag cgc agc gac    1009
Glu Ala Thr Ala Leu Ala Ala Gly Asp Ala Pro Val Glu Arg Ser Asp
            160                 165                 170

gcg gcc cgc atc cag gcg gcg gag gcg cgc gcc acg ggg ctg gac gcg    1057
Ala Ala Arg Ile Gln Ala Ala Glu Ala Arg Ala Thr Gly Leu Asp Ala
        175                 180                 185

aac gtg ccc ggc ggc ctg gcc cgg cag gcg cag tcg gcc gcg gcg gcc    1105
```

```
Asn Val Pro Gly Gly Leu Ala Arg Gln Ala Gln Ser Ala Ala Ala Ala
    190                 195                 200

aac tcg tgg gcg tgg gga gac gag gac aag gcc acg ctc ggc gac gtc      1153
Asn Ser Trp Ala Trp Gly Asp Glu Asp Lys Ala Thr Leu Gly Asp Val
205                 210                 215                 220

ctg gcg gtacgagtca cgaacacgac gtgccatcgt tttcgtttcg tgccgctgct       1209
Leu Ala


gctatatatc tgacagtgcg tgttggtggt gcaacagagc agagatcttt tgactatttg    1269

ttctttgtcg tacgtgcag aac gcg acg gcg agg ttg gtg gcg gac aag ccg     1321
                     Asn Ala Thr Ala Arg Leu Val Ala Asp Lys Pro
                             225                 230

gtg gag agc gcc gat gcg ttg ggg gtg gct ggc gcg gag aac cgc aac      1369
Val Glu Ser Ala Asp Ala Leu Gly Val Ala Gly Ala Glu Asn Arg Asn
    235                 240                 245

agg aac gac ggg acg gcg agg ccc gga ggc gtg gcg gcg tcc atg gct      1417
Arg Asn Asp Gly Thr Ala Arg Pro Gly Gly Val Ala Ala Ser Met Ala
250                 255                 260                 265

gcg gcc gca cgg ctc aac cgt aac gag gcg gtc tgg gag tgaagcagct       1466
Ala Ala Ala Arg Leu Asn Arg Asn Glu Ala Val Trp Glu
            270                 275

gcctggagag gagacacgtg cgtgtcctgg actctgaagt cctcgtcttt ttttttgttc    1526

gctagctagc tctgtacctc agcgcacgct ttacctacgt ccattcaggc gatcgagctg    1586

tgtaaatatg tagtatgtga cggctcagaa cgtgtcagtg tgtgtaactc gacatcaggc    1646

gatcgagctg tgtaaatatg tagtgttgta ccttcgtgca atataataaa gtaagatacg    1706

cgcgcgtcaa aagcgtgacc ggtgtaagat atactccgta tgcacataat taaggtgcat    1766

gctgtatatg gtgaagatgt gttaccccag tcgtacaaac caatatattt agccgggcgt    1826

gcggcgtgct atagtcgggt                                                1846


&lt;210&gt;  32
&lt;211&gt;  278
&lt;212&gt;  PRT
&lt;213&gt;  Zea mays

&lt;400&gt;  32

Met Ser Gln Glu Gln Pro Arg Arg Pro Ser Gly His Glu Glu Thr Ser
1               5                   10                  15


Gly Gly Gly Glu Gln Gly Ala Val Arg Tyr Gly Asp Val Phe Pro Ala
            20                  25                  30


Val Ser Gly Gly Leu Ala Glu Lys Pro Val Ala Arg Arg Thr Ala Thr
        35                  40                  45
```

```
Met Gln Ser Ala Glu Asn Leu Val Phe Gly Gln Thr Leu Lys Gly Gly
    50                  55                  60

Pro Ala Ala Ala Met Gln Ser Ala Ala Thr Thr Asn Glu Arg Met Gly
65                  70                  75                  80

Ala Val Gly His Asp Gln Ala Thr Asp Ala Thr Ala Val Gln Gly Val
                85                  90                  95

Thr Val Ser Glu Thr Arg Val Pro Gly Gly Gly Arg Ile Val Thr Glu
            100                 105                 110

Phe Val Ala Gly Gln Ala Val Gly Gln Tyr Leu Ala Arg Asp Asp Asp
        115                 120                 125

Gly Gly Gly Gly Ile Ala Gly Pro Gly Ala Gly Ala Gly Val Ala Gly
    130                 135                 140

Lys Asp Ile Thr Lys Val Thr Ile Gly Glu Ala Leu Glu Ala Thr Ala
145                 150                 155                 160

Leu Ala Ala Gly Asp Ala Pro Val Glu Arg Ser Asp Ala Ala Arg Ile
                165                 170                 175

Gln Ala Ala Glu Ala Arg Ala Thr Gly Leu Asp Ala Asn Val Pro Gly
            180                 185                 190

Gly Leu Ala Arg Gln Ala Gln Ser Ala Ala Ala Ala Asn Ser Trp Ala
        195                 200                 205

Trp Gly Asp Glu Asp Lys Ala Thr Leu Gly Asp Val Leu Ala Asn Ala
    210                 215                 220

Thr Ala Arg Leu Val Ala Asp Lys Pro Val Glu Ser Ala Asp Ala Leu
225                 230                 235                 240

Gly Val Ala Gly Ala Glu Asn Arg Asn Arg Asn Asp Gly Thr Ala Arg
                245                 250                 255

Pro Gly Gly Val Ala Ala Ser Met Ala Ala Ala Ala Arg Leu Asn Arg
            260                 265                 270

Asn Glu Ala Val Trp Glu
        275
```

<210> 33
<211> 501
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (1)..(501)
<223> DEHYDRIN DHN1 (RAB-17 PROTEIN)

<400> 33
atg gag tac ggt cag cag ggg cag cac ggc cac ggc gcc acg ggc cat 48
Met Glu Tyr Gly Gln Gln Gly Gln His Gly His Gly Ala Thr Gly His
1 5 10 15

gtc gac cag tac ggc aac cca gtc ggc ggc gtc gag cac ggc acc ggc 96
Val Asp Gln Tyr Gly Asn Pro Val Gly Gly Val Glu His Gly Thr Gly
20 25 30

ggc atg agg cac ggc acg gga acc ggc ggc atg ggc cag ctg ggt gag 144
Gly Met Arg His Gly Thr Gly Thr Gly Gly Met Gly Gln Leu Gly Glu
35 40 45

cac ggc ggc gct ggc atg ggt ggc ggg cag ttc cag cct gcg agg gag 192
His Gly Gly Ala Gly Met Gly Gly Gly Gln Phe Gln Pro Ala Arg Glu
50 55 60

gag cac aag acc ggc ggc atc ctg cat cgc tcc ggc agc tcc agc tcc 240
Glu His Lys Thr Gly Gly Ile Leu His Arg Ser Gly Ser Ser Ser Ser
65 70 75 80

agc tcg tcg gag gac gac ggc atg ggc gga agg agg aag aag gga atc 288
Ser Ser Ser Glu Asp Asp Gly Met Gly Gly Arg Arg Lys Lys Gly Ile
85 90 95

aag gag aag atc aag gag aag ctg ccc gga ggc cac aag gac gac cag 336
Lys Glu Lys Ile Lys Glu Lys Leu Pro Gly Gly His Lys Asp Asp Gln
100 105 110

cac gcc acg gcg acg acc ggc ggc gcc tac ggg cag cag gga cac acc 384
His Ala Thr Ala Thr Thr Gly Gly Ala Tyr Gly Gln Gln Gly His Thr
115 120 125

ggc agc gcc tac ggg cag cag gga cac acc ggc ggc gcc tac gcc acc 432
Gly Ser Ala Tyr Gly Gln Gln Gly His Thr Gly Gly Ala Tyr Ala Thr
130 135 140

ggc acc gag ggc acc ggc gag aag aaa ggc att atg gac aag atc aaa 480
Gly Thr Glu Gly Thr Gly Glu Lys Lys Gly Ile Met Asp Lys Ile Lys
145 150 155 160

gag aag ctg ccc gga cag cac 501
Glu Lys Leu Pro Gly Gln His
165

<210> 34
<211> 167
<212> PRT
<213> Zea mays

<400> 34

Met Glu Tyr Gly Gln Gln Gly Gln His Gly His Gly Ala Thr Gly His
1 5 10 15

Val Asp Gln Tyr Gly Asn Pro Val Gly Gly Val Glu His Gly Thr Gly
20 25 30

Gly Met Arg His Gly Thr Gly Thr Gly Gly Met Gly Gln Leu Gly Glu
35 40 45

His Gly Gly Ala Gly Met Gly Gly Gly Gln Phe Gln Pro Ala Arg Glu
50 55 60

Glu His Lys Thr Gly Gly Ile Leu His Arg Ser Gly Ser Ser Ser Ser
65 70 75 80

Ser Ser Ser Glu Asp Asp Gly Met Gly Gly Arg Arg Lys Lys Gly Ile
85 90 95

Lys Glu Lys Ile Lys Glu Lys Leu Pro Gly Gly His Lys Asp Asp Gln
100 105 110

His Ala Thr Ala Thr Thr Gly Gly Ala Tyr Gly Gln Gln Gly His Thr
115 120 125

Gly Ser Ala Tyr Gly Gln Gln Gly His Thr Gly Gly Ala Tyr Ala Thr
130 135 140

Gly Thr Glu Gly Thr Gly Glu Lys Lys Gly Ile Met Asp Lys Ile Lys
145 150 155 160

Glu Lys Leu Pro Gly Gln His
165

<210> 35

<211> 1243

<212> DNA

<213> wheat

<220>

<221> CDS

<222> (51)..(1226)

<223> Dehydrin

<400> 35

```
gtaaacacat cagcactagt agatttcacg agtcagaagc tcagcgcaag atg gag        56
                                                       Met Glu
                                                       1

aac cag gca cac atc gcc ggc gag aag aag ggc atc atg gag aag atc      104
Asn Gln Ala His Ile Ala Gly Glu Lys Lys Gly Ile Met Glu Lys Ile
        5                   10                  15

aag gag aag ctc ccc ggc ggc cac ggc gac cac aag gag acc gct ggt      152
Lys Glu Lys Leu Pro Gly Gly His Gly Asp His Lys Glu Thr Ala Gly
    20                  25                  30

acc cac ggg cac gcc gcc acg gcg acg cat ggt gcc ccg gcc acc ggt      200
Thr His Gly His Ala Ala Thr Ala Thr His Gly Ala Pro Ala Thr Gly
35                  40                  45                  50

ggt gcc tac ggg cag cag ggt cac gct gga acc acc ggc acg ggg ttg      248
Gly Ala Tyr Gly Gln Gln Gly His Ala Gly Thr Thr Gly Thr Gly Leu
                55                  60                  65

cat ggc gcc cac gcc ggc gag aag aag ggc gtg atg gag aac atc aag      296
His Gly Ala His Ala Gly Glu Lys Lys Gly Val Met Glu Asn Ile Lys
            70                  75                  80

gac aag ctc cct ggt ggc cac gag gac cac cag cag acc ggt ggc cac      344
Asp Lys Leu Pro Gly Gly His Glu Asp His Gln Gln Thr Gly Gly His
        85                  90                  95

tac ggg cag cag gga cac gcc ggc acg gcg acg cat ggc acc ccg gct      392
Tyr Gly Gln Gln Gly His Ala Gly Thr Ala Thr His Gly Thr Pro Ala
    100                 105                 110

acc gct ggc acc tat ggg caa cag ggg cat acc ggc acg gcg acg cat      440
Thr Ala Gly Thr Tyr Gly Gln Gln Gly His Thr Gly Thr Ala Thr His
115                 120                 125                 130

ggc acc cca gcg acc ggt ggc acc tat ggg gag cag gga cac acc gga      488
Gly Thr Pro Ala Thr Gly Gly Thr Tyr Gly Glu Gln Gly His Thr Gly
                135                 140                 145

gtg acc ggc acg ggg acg cac ggc acc ggc gag aag aag ggc ctc atg      536
Val Thr Gly Thr Gly Thr His Gly Thr Gly Glu Lys Lys Gly Leu Met
            150                 155                 160

gag aac atc aag gag aag ctc cct ggt ggc cat ggt gac cac cag cag      584
Glu Asn Ile Lys Glu Lys Leu Pro Gly Gly His Gly Asp His Gln Gln
        165                 170                 175

acc gct ggc acc tac ggg cag cag gga cac gtc ggc acg ggg aca cat      632
Thr Ala Gly Thr Tyr Gly Gln Gln Gly His Val Gly Thr Gly Thr His
    180                 185                 190

ggc gcc ccg gct acc ggc ggg gcc tac ggg cag cat gaa cac gcc gga      680
Gly Ala Pro Ala Thr Gly Gly Ala Tyr Gly Gln His Glu His Ala Gly
```

```
195                 200                 205                 210

gtg gcc ggc gcg gga aca tac ggc acc ggc gag aag aag ggc gtc atg       728
Val Ala Gly Ala Gly Thr Tyr Gly Thr Gly Glu Lys Lys Gly Val Met
                215                 220                 225

gag aac atc aag gac aag ctc cct ggc ggc cac ggc gac cac cag cag       776
Glu Asn Ile Lys Asp Lys Leu Pro Gly Gly His Gly Asp His Gln Gln
            230                 235                 240

acc ggt ggc acc tac ggg cag cag gga cac acc ggc acg gcg acg cat       824
Thr Gly Gly Thr Tyr Gly Gln Gln Gly His Thr Gly Thr Ala Thr His
        245                 250                 255

ggc acc ccg gcc ggc ggc ggc acc tat gag cag cac gga cac acc ggg       872
Gly Thr Pro Ala Gly Gly Gly Thr Tyr Glu Gln His Gly His Thr Gly
    260                 265                 270

atg acc ggc acg ggg aca cac ggc acc ggc gag aag aag ggc gtc atg       920
Met Thr Gly Thr Gly Thr His Gly Thr Gly Glu Lys Lys Gly Val Met
275                 280                 285                 290

gag aac atc aag gag aag ctc ccc ggt ggc cac ggc gac cac cag cag       968
Glu Asn Ile Lys Glu Lys Leu Pro Gly Gly His Gly Asp His Gln Gln
                295                 300                 305

acc ggt gga gcc tac ggg cag cag gga cac acc ggc acg gcg acg cat      1016
Thr Gly Gly Ala Tyr Gly Gln Gln Gly His Thr Gly Thr Ala Thr His
            310                 315                 320

ggc act ccg gct ggc ggc ggc acc tac ggg cag cat gca cac act gga      1064
Gly Thr Pro Ala Gly Gly Gly Thr Tyr Gly Gln His Ala His Thr Gly
        325                 330                 335

atg acc ggc acg gag acg cac ggc acc acg gcc acc ggc ggc acc cat      1112
Met Thr Gly Thr Glu Thr His Gly Thr Thr Ala Thr Gly Gly Thr His
    340                 345                 350

ggg cag cac gga cac gcc gga acg act ggc act ggg aca cac ggc acc      1160
Gly Gln His Gly His Ala Gly Thr Thr Gly Thr Gly Thr His Gly Thr
355                 360                 365                 370

gac ggg gtg ggc gag aag aag agc ctc atg gac aag atc aag gac aag      1208
Asp Gly Val Gly Glu Lys Lys Ser Leu Met Asp Lys Ile Lys Asp Lys
                375                 380                 385

ctg cct gga cag cac tga gcccggtgtg ccgacgg                           1243
Leu Pro Gly Gln His
            390
```

<210> 36

<211> 391

<212> PRT

<213> wheat

<400> 36

Met Glu Asn Gln Ala His Ile Ala Gly Glu Lys Lys Gly Ile Met Glu
1 5 10 15

Lys Ile Lys Glu Lys Leu Pro Gly Gly His Gly Asp His Lys Glu Thr
20 25 30

Ala Gly Thr His Gly His Ala Ala Thr Ala Thr His Gly Ala Pro Ala
35 40 45

Thr Gly Gly Ala Tyr Gly Gln Gln Gly His Ala Gly Thr Thr Gly Thr
50 55 60

Gly Leu His Gly Ala His Ala Gly Glu Lys Lys Gly Val Met Glu Asn
65 70 75 80

Ile Lys Asp Lys Leu Pro Gly Gly His Glu Asp His Gln Gln Thr Gly
85 90 95

Gly His Tyr Gly Gln Gln Gly His Ala Gly Thr Ala Thr His Gly Thr
100 105 110

Pro Ala Thr Ala Gly Thr Tyr Gly Gln Gln Gly His Thr Gly Thr Ala
115 120 125

Thr His Gly Thr Pro Ala Thr Gly Gly Thr Tyr Gly Glu Gln Gly His
130 135 140

Thr Gly Val Thr Gly Thr Gly Thr His Gly Thr Gly Glu Lys Lys Gly
145 150 155 160

Leu Met Glu Asn Ile Lys Glu Lys Leu Pro Gly Gly His Gly Asp His
165 170 175

Gln Gln Thr Ala Gly Thr Tyr Gly Gln Gln Gly His Val Gly Thr Gly
180 185 190

Thr His Gly Ala Pro Ala Thr Gly Gly Ala Tyr Gly Gln His Glu His
195 200 205

Ala Gly Val Ala Gly Ala Gly Thr Tyr Gly Thr Gly Glu Lys Lys Gly
210 215 220

Val Met Glu Asn Ile Lys Asp Lys Leu Pro Gly Gly His Gly Asp His
225 230 235 240

Gln Gln Thr Gly Gly Thr Tyr Gly Gln Gln Gly His Thr Gly Thr Ala
245 250 255

Thr His Gly Thr Pro Ala Gly Gly Gly Thr Tyr Glu Gln His Gly His
260 265 270

Thr Gly Met Thr Gly Thr Gly Thr His Gly Thr Gly Glu Lys Lys Gly
275 280 285

Val Met Glu Asn Ile Lys Glu Lys Leu Pro Gly Gly His Gly Asp His
290 295 300

Gln Gln Thr Gly Gly Ala Tyr Gly Gln Gln Gly His Thr Gly Thr Ala
305 310 315 320

Thr His Gly Thr Pro Ala Gly Gly Gly Thr Tyr Gly Gln His Ala His
325 330 335

Thr Gly Met Thr Gly Thr Glu Thr His Gly Thr Thr Ala Thr Gly Gly
340 345 350

Thr His Gly Gln His Gly His Ala Gly Thr Thr Gly Thr Gly Thr His
355 360 365

Gly Thr Asp Gly Val Gly Glu Lys Lys Ser Leu Met Asp Lys Ile Lys
370 375 380

Asp Lys Leu Pro Gly Gln His
385 390

<210> 37
<211> 5
<212> PRT
<213> Zea mays

<400> 37

Ile Ser Tyr Glu Leu
1 5

<210> 38
<211> 15
<212> PRT
<213> Zea mays

<400> 38

Phe Asn Asn Lys Val Phe Cys Leu Met Phe Val Ala Ser Gln Lys
1 5 10 15

<210> 39
<211> 6
<212> PRT
<213> Zea mays

<400> 39

Glu Ser Thr Phe Leu Asp
1 5

<210> 40
<211>
<212> PRT
<213> Zea mays

<400> 40

<400> 3

Pro Gln Cys Ser Gln Gln Tyr Leu Ser Pro Val Thr Ala Ala Arg
1 5 10 15

<210> 41
<211>
<212> PRT
<213> Zea mays

<400> 41

Pro Ser Ala Thr Ser Thr Asn Ser Glu Thr Ala Ala Phe Ala Ser Ala
1 5 10 15

Arg

<210> 42
<211>
<212> PRT
<213> barley

<400> 42

Lys Val Ala Leu Val Thr Gly Gly Asp Ser Gly Ile Gly Arg
1 5 10

<210> 43
<211>
<212> PRT
<213> Zea mays

<400> 43

Lys Gly Leu Ala Tyr Glu Tyr Leu Glu Gln Asp Leu Gly Asn Lys
1 5 10 15

<210> 44
<211>
<212> PRT
<213> Zea mays

<400> 44

Arg Pro Gly Gly Val Ala Ala Ser Met Ala Ala Ala Ala Arg
1 5 10

<210> 45
<211>
<212> PRT
<213> Zea mays

<400> 45

Thr Gly Gly Met Arg His Gly Thr Gly Thr Thr Gly Gly Met Gly Gln
1 5 10 15

Leu Gly Glu His Gly Gly Ala Gly Met Gly Gly Gly Gln Phe Gln Pro
20 25 30

Ala Arg